Digitized by Arya Samaj Foundation Chennai and eGangotri

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

I. PHIL

Vol.87
1990

G. K. V.
Hardwar

In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri
11659
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

110159

Digitized by Arya Samaj Foundation Chennai and eGangotri

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

51
L.
87

VOLUME LXXXVII, NUMBER 1
JANUARY 1990

1/M [illegible]

page

5 *The Division of Cognitive Labor* Philip Kitcher
23 *Aesthetic Qualities and Aesthetic Value* Alan H. Goldman

BOOK REVIEWS
38 Kent Bach: *Thought and Reference* William W. Taschek

46 MEMO TO AUTHORS
47 NOTE TO CONTRIBUTORS & STATEMENT OF COPYRIGHT POLICY
48 NEW BOOKS

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Charles Larmore, Isaac Levi, Mary Mothersill, Charles D. Parsons, James J. Walsh. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; *Other Trustees,* Leigh S. Cauman and André Schiffrin.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, New York 10027.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1990 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy,* 709 Philosophy Hall, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVII, NUMBER 2
FEBRUARY 1990

page

57 *In Defense of '⊃'* James F. Thomson

71 *Personhood and Personal Identity* Marya Schechtman

BOOK REVIEWS

93 Patricia Churchland: *Neurophilosophy: Toward a Unified Theory of the Mind/Brain* Kathleen A. Akins

103 Peter Galison: *How Experiments End* Ian Hacking

106 Philip Pettit and John McDowell, eds.: *Subject, Thought, and Context* Christopher S. Hill

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Charles Larmore, Isaac Levi, Mary Mothersill, Charles D. Parsons, James J. Walsh. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; *Other Trustees,* Leigh S. Cauman and André Schiffrin.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, New York 10027.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1990 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY and additional mailing offices. Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVII, NUMBER 3
MARCH 1990

page

113 *Descriptive Pronouns and Donkey Anaphora*
Stephen Neale

COMMENTS AND CRITICISM

151 *The Poverty of Pluralism: A Reply to Sterelny and Kitcher*
Elliott Sober

158 *The Illusory Riches of Sober's Monism*
Philip Kitcher, Kim Sterelny, and C. Kenneth Waters

BOOK REVIEWS

161 Marjorie Grene: *Descartes* — Lynn S. Joy

165 NOTES AND NEWS

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Charles Larmore, Isaac Levi, Mary Mothersill, Charles D. Parsons, James J. Walsh. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; *Other Trustees,* Leigh S. Cauman and André Schiffrin.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, New York 10027.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1990 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy,* Copy, 709 Philosophy, Columbia U, NYC 10027.

CC-0. In Public Domain. Panini Kanya Maha Vidyalaya Collection. Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVII, NUMBER 4
APRIL 1990

page

169 *The Reward Event and Motivation* Carolyn R. Morillo

187 *Mutual Benevolence and the Theory of Happiness* David M. Estlund

COMMENTS AND CRITICISM

205 *A Second Look at Pornography and the Subordination of Women* W. A. Parent

BOOK REVIEWS

212 Daniel Dennett: *The Intentional Stance* Sydney Shoemaker

217 NEW BOOKS

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Charles Larmore, Isaac Levi, Mary Mothersill, Charles D. Parsons, James J. Walsh. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; *Other Trustees,* Leigh S. Cauman and André Schiffrin.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, New York 10027.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1990 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVII, NUMBER 5
MAY 1990

141200

page

225 *Second Thoughts about Church's Thesis and Mathematical Proofs* Elliott Mendelson

234 *Second-order Logic, Foundations, and Rules* Stewart Shapiro

COMMENTS AND CRITICISM

262 *Sensationalized Philosophy: A Reply to Marquis's "Why Abortion is Immoral"* Ann E. Cudd

264 *Does a Fetus Already Have a Future-like-ours?* Peter K. McInerney

268 *Killing, Abortion, and Contraception: A Reply to Marquis* Alastair Norcross

278 NOTES AND NEWS

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Charles Larmore, Isaac Levi, Mary Mothersill, Charles D. Parsons, James J. Walsh. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; *Other Trustees,* Leigh S. Cauman and André Schiffrin.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, New York 10027.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1990 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy,* Journal of Philosophy, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVII, NUMBER 6
JUNE 1990

page

THE DEWEY LECTURES 1989:
279 *The Structure and Content of Truth* Donald Davidson

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill, James J. Walsh. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; *Other Trustees,* Leigh S. Cauman, Kent Greenawalt, André Schiffrin.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, New York 10027.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1990 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027.

CC-0. In Public Domain. Panini Kanya Maha Vidyalaya Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

31/8/90

VOLUME LXXXVII, NUMBER 7
JULY 1990

page

329 *Utilitarianism, Sociobiology, and the Limits of Benevolence* Danny Scoccia

346 *Commodities, Language, and Desire* Michael Bacharach

COMMENTS AND CRITICISM

369 *Misunderstandings of Epistemic TIT for TAT: Reply to John Woods* Michel J. Blais

BOOK REVIEWS

375 Brian Barry: *A Treatise of Social Justice, Vol. I: Theories of Justice* Thomas W. Pogge

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill, James J. Walsh. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; *Other Trustees,* Leigh S. Cauman, Kent Greenawalt, André Schiffrin.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, New York 10027.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1990 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY and additional mailing offices. Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVII, NUMBER 8
AUGUST 1990

page

385 *Identity Statements and the Necessary A Posteriori*
Helen Steward

399 *What Is Reasoning? What Is an Argument?*
Douglas N. Walton

COMMENTS AND CRITICISM

420 *Virtue and Ignorance* Owen Flanagan

BOOK REVIEWS

429 Edward Craig: *The Mind of God and the Works of Man*
Stephen Gaukroger

432 Norman Swartz: *The Concept of Physical Law*
Lawrence Sklar

436 NOTES AND NEWS

437 NEW BOOKS: TRANSLATIONS

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill, James J. Walsh. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; *Other Trustees,* Leigh S. Cauman, Kent Greenawalt, André Schiffrin.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, New York 10027.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1990 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to: *Journal of Philosophy*, 709 Philosophy, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVII, NUMBER 9
SEPTEMBER 1990

page
439 *National Self-determination* Avishai Margalit & Joseph Raz
462 *Troubles with Ockhamism* David Widerker

COMMENTS AND CRITICISM
481 *Pareto Unanimity and Consensus* Isaac Levi

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill, James J. Walsh. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; *Other Trustees,* Leigh S. Cauman, Kent Greenawalt, André Schiffrin.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, New York 10027.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1990 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to Journal of Philosophy, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVII, NUMBER 10
OCTOBER 1990

Special Issue, Eighty-seventh Annual Meeting
American Philosophical Association, Eastern Division

page

Symposium: Gender, Race, Ethnicity: Sources of Identity

493 *"But Would That Still Be Me?" Notes on Gender, "Race," Ethnicity, as Sources of "Identity"* Anthony Appiah

500 *Structure/Antistructure and Agency under Oppression* Maria C. Lugones

508 *Comments on Appiah and Lugones* Thomas E. Wartenberg

Symposium: John Locke after Three-hundred Years

510 *The Way of Ideas: A Retrospective* John W. Yolton

517 *Political Freedom* James Tully

524 *Locke's Moral Psychology* Vere Chappell

Symposium: Philosophy of Sidney Hook

526 *Pragmatic Naturalism: The Philosophy of Sidney Hook (1902–1989)* Paul Kurtz

535 *Hook, Dewey, and Marx* Nicholas Capaldi

Symposium: Philosophy of Wilfrid Sellars

537 *Sellars's Two Images of the World* Bruce Aune

546 *Response to Aune* Jay F. Rosenberg

548 Johnsonian Prize Announcement

549 NOTES AND NEWS

551 Program of the Meetings

564 List of Program Participants

569 Appendix: Programs of Group Meetings

580 List of Group Participants

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill, James J. Walsh. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; *Other Trustees,* Leigh S. Cauman, Kent Greenawalt, André Schiffrin.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, New York 10027.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1990 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027.

CC-0. In Public Domain. Panini Kanya Maha Vidyalaya Collection.

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVII, NUMBER 11
NOVEMBER 1990

Special Issue, Eighty-seventh Annual Meeting
American Philosophical Association, Eastern Division

page

Symposium: Centennial of William James's *Principles of Psychology*

585 *The Pragmatist Sieve of Concepts: Description versus Interpretation* Charlene Haddock Seigfried

593 *James and Freud* Gerald E. Myers

600 *Comments on Seigfried and Myers* H. S. Thayer

Symposium: Moral Cognitivism and Realism

602 *Meaning and Value* Stephen Schiffer

615 *Schiffer on Meaning and Value* Nicholas L. Sturgeon

617 *A Relativist Alternative to Antirealism* David B. Wong

Symposium: Realism and Philosophy of Mathematics

619 *A Problem in the Foundations of Set Theory* Penelope Maddy

Symposium: African Famine

629 *African Famine: New Economic and Ethical Perspectives* George R. Lucas, Jr.

642 *Famine and Distribution* William Aiken

Symposium: The Vanishing Subject of Contemporary Discourse

644 *The Vanishing Subject of Contemporary Discourse: A Pragmatic Response* Vincent M. Colapietro

656 *Subjects Constructed, Deconstructed, and Reconstructed* John J. Stuhr

Symposium: Cartesian Skepticism

658 *Cartesian Skepticism and Inference to the Best Explanation* Jonathan Vogel

667 *Is Realism Really the Best Hypothesis?* Berent Enç

669 NOTES AND NEWS

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill, James J. Walsh. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; *Other Trustees,* Leigh S. Cauman, Kent Greenawalt, André Schiffrin.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, New York 10027.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1990 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy*, 709 Philosophy Hall, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVII, NUMBER 12
DECEMBER 1990

page

671 *Truth and Language-world Connections* Dorothy Grover

688 *Rights and Rights Violators: A New Approach to the Nature of Rights* Heather J. Gert

695 *Supervenience Is a Two-way Street* Richard B. Miller

BOOK REVIEWS

702 Mary Mothersill: *Beauty Restored* Ted Cohen

708 Richard Rorty: *Contingency, Irony, and Solidarity* Alasdair MacIntyre

711 Nelson Goodman and Catherine Z. Elgin: *Reconceptions in Philosophy and Other Arts and Sciences* Jonathan Adler

716 Thomas W. Pogge: *Realizing Rawls* Robert Paul Wolff

721 NEW BOOKS

730 INDEX TO VOLUME LXXXVII

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill, James J. Walsh. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; *Other Trustees,* Leigh S. Cauman, Kent Greenawalt, André Schiffrin.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, New York 10027.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1990 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy,* 709 Philosophy Hall, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVII
JANUARY–DECEMBER 1990

Edited by
Bernard Berofsky, Arthur C. Danto
Charles Larmore, Isaac Levi
Sidney Morgenbesser, Mary Mothersill
Charles D. Parsons, James J. Walsh

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1990 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy*, 709 Philosophy, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVII, NO. 1, JANUARY 1990

THE DIVISION OF COGNITIVE LABOR*

BY 1804, the phlogiston theory was dead. Thirty years earlier, the same theory had been favored by almost every chemist in Europe. If the chemical revolution was resolved on the basis of reason and evidence, then it appears that there must have been some moment between 1774 and 1804 when the balance of evidence finally tipped against the phlogiston theory and in favor of Antoine-Laurent Lavoisier's "new chemistry."

Imagine that the objective degree of confirmation of the phlogiston theory just prior to noon on April 23, 1787, was 0.51, that of the new chemistry 0.49. At noon, Lavoisier performed an important experiment, and the degrees of confirmation shifted to 0.49 and 0.51, respectively. Allowing for a time lag in the dissemination of the critical information, we can envisage that there was a relatively short interval after noon on April 23, 1787, before which all rational chemists were phlogistonians, and after which all were followers of Lavoisier.

Does this scenario of initially uniform opinion, sudden jumping of ship, and new consensus signal the rational growth of scientific knowledge? If you had been a philosopher-monarch, concerned to have your scientist-subjects distribute their efforts so as to promote the eventual attainment of truth by the community, you would (rightly) have dismissed this assignment of resources (the scientists themselves) as a bad bargain. With the evidential balance between

* I am grateful to the many people who have heard or read ancestors of this paper and who have given me valuable advice, and, in particular, to John Beatty, Gerald Doppelt, Isaac Levi, David Lewis, and Elisabeth Lloyd. I owe a large debt to the ideas and the writings of Thomas Kuhn. Special thanks are due to Stephen Stich for his many detailed comments and constructive suggestions.

0022-362X/90/8701/5–22 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the two theories so delicate, you would have preferred that some scientists were not quite so clear-headed in perceiving the merits of the theories, so that the time of uniform decision was postponed. A community of chemists that responded in the fashion of my original story is a badly-run community—an irrational community, if you like.

My story is intended to raise an important, if neglected,[1] problem about the growth of science. Is it possible that there should be a mismatch between the demands of individual rationality and those of collective (or community) rationality? Could it turn out that high-minded inquirers, following principles of individual rationality, should do a poor job of promoting the epistemic projects of the community that they constitute? Might those with baser motives actually do more to advance their community's epistemic endeavors? Are there conditions under which, in light of our goals as an epistemic community, we ought to want to maintain cognitive diversity? What, if anything, do we do, or can we do, about it?

I

Perhaps the predicament I describe is simply an artifact of faulty presuppositions. The following brief remarks are intended to address some obvious concerns and to indicate that the problems I have raised survive modifications of the framework I have employed in presenting them.

Any optimistic suggestion that the kind of delicate balance I have envisaged is unrealistic because scientific decisions are always clear-cut is belied by recent studies in the history of science. Moreover, even where there are large differences in measures of support, it is still possible for discrepancies between individual and collective rationality to arise. A more promising diagnosis of what goes wrong in my fable maintains that we cannot speak sensibly about numerical values representing the empirical support a theory enjoys. The more limited version of this tactic allows that degrees of support should be understood as connected subintervals of [0,1]. But it is easy to construct versions of my original story in which we would prefer a mi-

[1] The problem is posed by Thomas Kuhn in "Objectivity, Value Judgment and Theory Choice" [in *The Essential Tension* (Chicago: University Press, 1977), pp. 320–339]. Virtually the only subsequent treatments of the possibility of discrepancies between individual and collective rationality are the proposals of Husain Sarkar in *A Theory of Method* (Berkeley: California UP, 1983), and my own very sketchy remarks in ch. 6 of *Abusing Science* (Cambridge: MIT, 1982). I have learned much from Sarkar's treatment of the topic, even though his focus is on alternatives in methodology rather than in differences in theories, research programs, or methods.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

nority of scientists to continue to espouse a theory whose objective support is measured by a subinterval of [0,0.5].

A more serious worry starts with the denial that there are objective measures of support. Does the sophisticated work in history of science not reveal to us that there are numerous cases in which equally reasonable people may disagree about the merits of rival theories, perhaps because they have different ideas about the significance of different problems or about the appropriate criteria for solving those problems? For the purposes of this essay, I do not want to take sides on this vexed question. I claim, simply, that we sometimes want to maintain cognitive diversity even in instances where it would be reasonable for all to agree that one of two theories was inferior to its rival, and we may be grateful to the stubborn minority who continue to advocate problematic ideas.

By the 1790s, only a handful of chemists continued to explore the possibility of reviving the phlogiston theory. As Thomas Kuhn[2] remarks, it was probably unreasonable for Joseph Priestley to persist as long as he did. Yet, from the point of view of the community of chemists, it was no bad thing that Priestley (and a few others) gave the phlogiston theory every last chance. Turning to our century, and to the history of Alfred Wegener's theory of continental drift, we can appreciate how things might have gone differently. In the 1920s and 1930s, Wegener's claim seemed to face insuperable difficulties, for there were apparently rigorous geophysical demonstrations that the forces required to move the continents would be impossibly large. Despite this, a few geologists, most notably Alexander du Toit, continued to advocate and articulate Wegener's ideas. I suggest that the distribution of cognitive effort was preferable to a situation in which even the small minority abandoned continental drift.[3]

Was it equally reasonable to be a drifter or an anti-drifter in the 1920s and 1930s? Inspired by appreciation of the intricate shifts in standards of appraisal that occur in the history of science, you might say, "yes." But then you face a problem of maintaining cognitive diversity of the same type as that with which I began: from the community's point of view, it would have been better if the geologists

[2] *The Structure of Scientific Revolutions* (Chicago: University Press, 1962; 2nd ed. 1970), p. 159.

[3] For an overview of the career of Wegener's theory, see Anthony Hallam, *A Revolution in the Earth Sciences* (New York: Oxford, 1973). Du Toit's persistence in supporting the theory is manifest in his book, *Our Wandering Continents* (Edinburgh: Oliver & Boyd, 1937).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

had been more equally divided. On the other hand, if you accept the idea that the geophysical arguments really did expose the implausibility of Wegener's theory, then the actual distribution of cognitive effort appears better—even though, of course, some of the epistemic agents, such as du Toit, are viewed as making an irrational choice.

Consider a last suggestion for avoiding puzzles about the division of cognitive labor. We can surely distinguish attitudes that scientists adopt toward theories, hypotheses, research programs, and so forth. In particular, we can differentiate *belief* in a theory from *pursuit* of research designed to apply or extend that theory.[4] Once we have recognized the distinction, can we not accept a simple solution to my puzzle? Whereas it may be rational for each of the scientists to believe the theory that is better supported by the available evidence, it may not be rational for each of them to pursue that theory, and what the community cares about is the distribution of pursuit not the distribution of belief.

This suggested way of avoiding the discrepancy between individual and collective rationality depends on adopting two principles of individual rationality, one for belief and one for pursuit. The idea that it is rational for a person to believe the better-supported theory seems, however, to be based on supposing that that person's aim is to achieve true beliefs (or some other desirable epistemic state, the acceptance of empirically adequate theories, for example). In that case, however, it appears that the person should also pursue the better-supported theory, since pursuing a doctrine that is likely to be false is likely to breed more falsehood (or less of the desired epistemic state). Only if we situate the individual in a society of other epistemic agents—as I shall try to do in later sections—does it begin to appear rational for someone to assign herself to the working out of ideas that she (and her colleagues) view as epistemically inferior.

II

Imagine, then, that you are a philosopher-monarch, with the prerogative of directing the course of scientific research. You hope to achieve certain epistemic goals—the construction of a complete, true story of the world, the articulation of an empirically adequate theory, the elimination of error, the solving of as many problems as possible, or whatever. You have an unerring eye for detecting the objective merits of theories and complete control of the scientific

[4] See Larry Laudan, *Progress and Its Problems* (Berkeley: California UP, 1977), pp. 108–114.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

workforce. What rule for the division of cognitive labor should you adopt?

Perhaps you decide to let your subjects be individually rational, allowing them to believe T whenever the objective epistemic merit of T is greater than that of its rivals (subject, perhaps, to a proviso that the merit of T be above some threshold value). But, as we saw, this is a poor strategy, liable to promote uniformity of opinion when you would prefer to keep your options open.

Even without the fiction of a philosopher-monarch, we can still consider the problem of the optimal community strategy for achieving epistemic ends. Continuing to be vague about what these ends are, I shall formulate the problem by distinguishing two types of epistemic intentions that individual scientists may have. X may have the intention that X may achieve some epistemic end (to whatever extent is possible): so, for example, X may intend that X acquire as many true beliefs as possible. This is X's *personal* epistemic intention. X may also have the intention that the community to which X belongs, the community of past, present, and future scientists, achieve an epistemic end (to whatever extent is possible): X may intend that the community of scientists uniformly adopt as many true theories as possible, in the long run. This is X's *impersonal* epistemic intention. We can recast the question "what is the rational community strategy"? as "how would scientists rationally decide to coordinate their efforts if their decisions were dominated by their impersonal epistemic intentions"? The fiction of the philosopher-monarch dramatizes the idea that this decision might require the subordination of personal epistemic intentions. If X is to engage in a community project with the goal that the community as a whole attain some epistemic end—an end that all X's fellow members also want to attain—then X may have to make decisions that do not coincide with those of an individually rational scientist. X should agree in advance that it may sometimes be necessary for some member(s) of the community to pursue (or even believe) an inferior theory, and that it may fall to X to play this role.[5]

I can now give a general description of the class of problems of optimal division of cognitive labor. Suppose that there is a set, S, of scientists, each of whom has a choice among the members of a set, R,

[5] Altruistically rational scientists are those who are prepared to pursue theories that they regard as inferior when, by doing so, they will promote achievement of the goals of their own (and their colleagues') impersonal epistemic intentions. Plainly this raises an even more bloodless ideal of scientific rationality than that criticized by historians and sociologists of science.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of rival cognitive objects. (R may be a set of rival theories, research programs, methods for approaching a problem, etc.) Each of the scientists has an impersonal epistemic intention that the community descending from S achieve some epistemic goal state: suppose that the intention of the ith scientist is that the community attain G_i (where the G_i may be different). For each of the scientists, there is an evaluation function, whose domain is R. The evaluation functions, which may be distinct, assess the epistemic merits of the members of R, and, I shall suppose, considerations of individual epistemic rationality dictate that each scientist adopt some cognitive attitude (appropriate to the category of objects belonging to R) in that member of R which ranks highest according to the scientist's evaluation function. (So, if the members of R are theories, the cognitive attitude may be belief, and the requirement of rationality may be that scientists believe that theory in R which comes out highest according to their evaluation function). The IR (individually rational) distribution of attitudes is that distribution generated among the members of S from the evaluation functions in accordance with the requirement of rationality. The CO (community optimum) distribution relative to i is that distribution of attitudes among the members of S which would maximize the probability of attaining G_i. There is a CO-IR discrepancy when there is a distribution of attitudes among the members of S which, for each i, yields a higher probability of attaining G_i than does the IR-distribution.

My general formulation allows for differences in scientists' impersonal intentions and differences in their assessments of epistemic merit, underscoring the point made in section I that problems about the division of cognitive labor arise even under the assumption that there may be changes in the standards of evaluation and changes in goals. The problems are easier to pose and easier to investigate, however, if we suppose uniformity in both respects: that is, that the G_i are all the same and that there is a single ("objective") evaluation function for each scientist. Just as I began by tacitly making this simplifying supposition, the rest of this essay will continue to adopt it. The problem only becomes more complicated if the supposition is discarded.

Given our simplification, we can talk about a community goal state, G, and an unrelativized CO-distribution. A CO-IR discrepancy will be a case in which the CO-distribution differs from the IR-distribution. I now want to descend from the abstract level of the last paragraphs, considering a particular type of problem of division of cognitive labor with the aim of identifying some conditions under

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

which CO-IR discrepancies can be expected to occur. Although my opening example concerned theory choice, the next two sections will explore the possibility of CO-IR discrepancies for the much more tractable case of choice between problem-solving methods. I shall return to the case of theory choice (which is algebraically more complex) in section V.

III

Once there was a very important molecule (VIM). Many people in the chemical community wanted to know the structure of VIM. Two methods for fathoming the structure were available. Method I involved using X-ray crystallography, inspecting the resultant photographs and using them to eliminate possibilities about bonding patterns. Method II involved guesswork and the building of tinker-toy models. Everybody agreed that the chances that an individual would discover the structure of VIM by using method I were greater than the chances that that individual would discover the structure by using method II. Since all members of the community were thoroughly rational, each chemist used method I. They are still working on the problem.

The community goal is to fathom the structure of VIM as quickly as possible. Suppose that each method is associated with a probability function, $p(n)$, representing the chance that the method will deliver an answer if n workers are assigned to it. Assume further that any answer delivered is recognizably either correct or incorrect. Imagine also that the relations between the probability functions represent their behavior over any time intervals we might consider—so, for any time interval, t, the probabilities that method I delivers an answer within t and that method II delivers an answer within t are in the same ratios as the functions p.[6] N workers are available for distribution between the two methods. The CO-distribution is given by having n workers use method I and $N - n$ use method II so as to maximize the probability that the structure of VIM will be discovered, that is, to maximize

$$p_1(n) + p_2(N - n) - Prob \text{ (both methods deliver)}$$

[6] These functions measure the chance that a method will deliver a correct answer, for an assignment of n workers, given that the world is as represented by the community's current knowledge about the molecule. Thus, for example, if little is known about VIM and if both methods have been pursued by similar numbers of workers for fathoming a large class of molecules, with method I proving successful much more frequently than method II, whatever the number of workers assigned, $p_1(n) > p_2(n)$. I shall not pursue here the question of exactly how to interpret the probabilities. To the best of my knowledge, it is possible to conceive of them along any of the currently popular lines.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

I shall assume, for simplicity's sake, that the probability that both methods will deliver the correct answer is zero. (The effect of this—nontrivial—assumption is solely to simplify the algebra. Qualitatively similar conclusions can be obtained, at far greater length, if it is not made.)

Evidently, the solution to the problem depends on the form of the functions $p(n)$. I shall call these *return functions*, since they measure the return in probability of reaching the goal for an investment of n workers. I shall take these functions to be subject to the following constraints: they should increase monotonically with n, they should be zero when n is zero, and they should tend asymptotically to some value p when n goes to infinity (p represents the intrinsic prospects of the method, the probability of its success when we abstract from limitations of human effort).[7] Given the simplifying assumption that both methods cannot work, we know that the values of the asymptotes, p_1 and p_2, must sum to less than 1. These constraints leave a lot of room for choice of functions. To make one point explicit, it is quite possible that the forms of the functions should be different for the two methods. (Imagine that one method responds much more quickly than the other to the efforts of workers.)

I shall consider two possibilities for the functions.[8] Suppose first that $p_i(n) = p_i(1 - e^{-kn})$. Then $p_1(n) + p_2(N - n)$ is maximized when

$$n = (kN + \ln p_1 - \ln p_2)/2k$$

Notice that, even when method I has more intrinsic promise than method II ($p_1 > p_2$), there is a range of conditions—when $\ln p_1 - \ln p_2 < kN$—under which the CO-distribution is to divide the community. Intuitively, a genuine division of cognitive labor would be best for the community if there is a large available workforce (N is large), or if the methods respond quickly to the injection of effort (k is not too small), or if the difference in intrinsic promise between the methods is not too great (p_1 and p_2 are fairly close). The inequality given above represents the ways in which tradeoffs are made.

[7] I originally thought that these constraints would apply in all cases. As Stephen Stich pointed out to me, however, too many cooks may spoil the broth. Imagine, for example, that the method involves observing some sensitive organisms and that crowding in the field would disturb the organisms' normal behavior.

[8] The two classes of functions I consider represent two major possibilities: either the rate of return is fast at the beginning, and then slows as the asymptote is approached, or the rate is initially slow, speeds up once a critical mass of workers has been assembled, and slows as saturation is reached.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The functions considered in the last paragraph have the property that the rate of increase in $p(n)$ is maximal when n is small. This idea may not be at all realistic. Perhaps the chances of achieving an answer by following a given method increase quite slowly at first, then go up rapidly once a critical mass of workers has accumulated, and eventually increase very slowly as saturation is approached. We can describe this behavior by mimicking the logistic-growth equation of population biology, supposing that the p_i are given by

$$p_i(n) = p_i(3n^2 - 2n^3/kN)/k^2N^2 \qquad (n < kN)$$
$$p_i(n) = p_i \qquad (n \geq kN)$$

If the probabilities are given by these functions, then there are various cases of interest, depending on the value of k. Provided that $k < \frac{1}{2}$, it is possible to realize the intrinsic prospects of both methods, so the CO-distribution divides the workforce. If $\frac{1}{2} < k < 1$, it is not hard to show that the optimal value of n is less than kN. When k is greater than or equal to 1, n should be N (recall that method I is superior, that is $p_1 > p_2$).

Let me give a qualitative interpretation of these findings. As in the previous case, k is a critical parameter, representing the responsiveness of the methods. If the methods are so responsive that the intrinsic prospects of both can be realized with the available workforce, then it is easy to appreciate that the community epistemic interests are best served by dividing the labor. Even when k is between $\frac{1}{2}$ and 1—so that it is possible to realize the intrinsic prospects of one method but not those of both—it may be better to divide the workforce so that the prospects of neither method are realized. Provided that the difference between p_1 and p_2 is not too great, it will be better to assign a new worker to method II, if method II already has sufficient devotees to offer a large return from a new investment, rather than to method I, if method I is nearly saturated. Once k reaches 1, however, it is always better to assign all resources to the method whose intrinsic prospects are higher.

Let us now turn to the IR-distribution for these cases. On one simple understanding of individual epistemic rationality, rational agents judge methods according to the intrinsic qualities of those methods, not according to what their fellows are doing. If we understand individual epistemic rationality in this simple way, then it is easy to see that there can be discrepancies between the IR-distribution ($\langle N, 0 \rangle$) and the CO-distribution (which is sometimes $\langle n, m \rangle$ with both n, m nonzero).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Perhaps we should think of individual epistemic rationality a bit differently. Suppose it is a requirement of individual epistemic rationality that an agent maximize her chances of following a method that yields the answer. We can interpret the requirement in two ways. (1) We imagine the agent making a decision in complete ignorance of what other members of the community are doing, so that the task is to choose i so that p_i (1) is as large as possible. (2) We imagine that the agent knows the current distribution $\langle r, s \rangle$, so that method I is to be chosen just in case $p_1(r+1) > p_2(s+1)$. On either interpretation, it is easy for CO-IR discrepancies to arise.

An obvious move at this point is to modify the requirements of individual epistemic rationality so that the discrepancies vanish by the magic of redefinition: simply declare that an individually rational agent is a person who chooses so as to belong to a community in which the chances of discovering the correct answer are maximized. I suggest that it is a virtue of the analysis I have been presenting that it forces into the open this altruistic ideal of rationality—an ideal that seems to me to be rather different from the concepts of rationality that figure in traditional philosophical discussions. But whether we identify community rationality and individual rationality by fiat, I am concerned with the properties of CO-distributions and the possibilities that real, imperfectly altruistic people might approximate them. The next section will explore the possibility that allowing our scientists to depart from the high-minded goals of individual rationality (and act on baser motives) might actually help the community's project.

IV

Plunging into some algebraic details, we left a community of chemists striving and failing to fathom VIM. I shall imagine that the CO-distribution for them involved a genuine division of labor (corresponding to one of those cases considered in the last section in which $p_1 > p_2, p_1(1) > p_2(1)$). They failed to achieve this, since all of them followed one of the principles of individual rationality which led to the distribution $\langle N, 0 \rangle$. Moreover, because the structure of VIM could only be fathomed by method II, their inability, as a community, to hedge their bets was costly.

By contrast, in a neighboring nation, the chemical community was composed of ruthless egoists. Each of the members of this community made decisions rationally, in the sense that actions were chosen to maximize the chances of achieving goals, but the goals were personal rather than epistemic. Those who elected to work on VIM did so because they believed that whoever discovered the structure of

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

VIM would win a much-coveted prize. Make the simplifying (but not altogether implausible) assumption that, if a method succeeds, then each person pursuing that method has an equal chance of winning the prize. How should we expect the Hobbesian community to distribute its effort?

Imagine that the community has reached a distribution $\langle n, N-n \rangle$. You are a scientist currently working on method I, and you ponder the possibility of switching to method II. The change would be good for you—given my assumption about your interests and aspirations—if it would increase the probability that you win the prize. Now the probability of your winning is the probability that someone in your group wins, divided by the number of group members. (Intuitively, by choosing a method, you buy into a lottery that has a probability of paying up, a probability dependent on the number of ticket holders; your chance of collecting anything is the probability that the lottery pays up divided by the number of tickets). Thus, at $\langle n, N-n \rangle$, it will behoove a scientist working on method I to switch to method II if

$$p_2(N-n+1)/(N-n+1) > p_1(n)/n$$

To understand how our imaginary Hobbesian agents might distribute themselves, we need to discover equilibria, points at which nobody is better off switching to the alternative method.[9] Let us say that the distribution $\langle n, N-n \rangle$ is *stable downward* if $p_1(n)/n$ is greater than or equal to $p_2(N-n+1)/(N-n+1)$, and *unstable downward* otherwise. Similarly, $\langle n, N-n \rangle$ is *stable upward* if $p_1(n+1)/(n+1)$ is less than or equal to $p_2(N-n)/(N-n)$, and *unstable upward* otherwise. $\langle n, N-n \rangle$ is *bilaterally stable* just in case it is both stable upward and stable downward.

If a community of Hobbesian scientists reaches a distribution that is bilaterally stable, then we can expect it to stay there. *Stability* is one thing, however, *attainability* another. Even though a particular distribution might be maintained, once it had been achieved, it may prove impossible for a group of self-interested scientists to reach it. For any distribution $\langle n, N-n \rangle$ that is bilaterally stable, we can define its *zone of attraction* to be the set of distributions that collapse to $\langle n, N-n \rangle$. More precisely, say that $\langle m, N-m \rangle$ collapses up to

[9] My thinking about the problem of the evolution of distributed effort in scientific communities has been heavily influenced by R. A. Fisher's classic discussion of the evolution of sex ratios [see his *The Genetical Theory of Natural Selection* (New York: Dover, 1958), pp. 158–160] and by the ideas of John Maynard Smith [particularly *Evolution and the Theory of Games* (New York: Cambridge, 1982)].

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

$\langle n, N-n \rangle$ just in case $m < n$, and for each x, $m < x < n$, $\langle x, N-x \rangle$ is unstable upward, and analogously for collapsing downward. $\langle n, N-n \rangle$ is *attainable* if its zone of attraction contains all distributions.

The Hobbesian community might work much better than the high-minded spirits of the last section who failed to divide the labor. More exactly, maybe there is a distribution which is both stable and attainable and which offers a higher probability of community success than the IR-distributions we considered above. (The ideal, of course, would be to show that the CO-distribution is both stable and attainable). The very factors that are frequently thought of as interfering with the rational pursuit of science—the thirst for fame and fortune, for example—might actually play a constructive role in our community epistemic projects, enabling us, as a group, to do far better than we would have done had we behaved like independent epistemically rational individuals. Or, to draw the moral a bit differently, social institutions within science might take advantage of our personal foibles to channel our efforts toward community goals rather than toward the epistemic ends that we might set for ourselves as individuals.

But is the possibility genuine? Consider cases. The simplest is that in which the return functions are given by

$$p_i(n) = p_i(1 - e^{-kn}) \text{ with } k \text{ large and } p_1 > p_2.^{10}$$

There is a bilaterally stable distribution in the neighborhood of $\langle n^*, N-n^* \rangle$, with $n^* = p_1 N/(p_1 + p_2)$. The distribution is attainable. Moreover, if p_1 is only slightly larger than p_2, the distribution yields a probability of community success that is close to that given by the CO-distribution. Moral: there are conditions under which the Hobbesians do better than their epistemically pure cousins, even conditions under which they come as close as you please to the ideal.

Life is more complicated if the return functions take the forms

$$p_i(n) = p_i(3n^2 - 2n^3/kN)/k^2N^2 \quad \text{for} \quad n < kN$$

$$p_i(n) = p_i \quad \text{for} \quad n > kN$$

$$\text{where} \quad p_2 < p_1 \quad \text{and} \quad k < p_2/(p_1 + p_2)$$

[10] It is also necessary for the claims that follow to be true that p_2 not be too small. These vague conditions can be formulated more precisely by requiring that $exp\{-kp_2N/(p_1 + p_2)\} << 1$.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Under these conditions, it is possible to achieve the intrinsic prospects of both methods and any distribution $\langle n, N - n \rangle$ with $kN < n < N - kN$ is a CO-distribution. There is a bilaterally stable distribution $\langle n^*, N - n^* \rangle$, given by $n^* = p_1 N/(p_1 + p_2)$. Provided that $p_2/(p_1 + p_2) > k$, this bilaterally stable distribution will be a CO-distribution. So far, so good. So long as the intrinsic prospects of the inferior method are not too low, and the methods respond quickly to the assignment of workers, there will be an optimal division of cognitive labor which the community can maintain—if it can but reach it.

But there is the rub. The zone of attraction of the stable distribution includes all the CO-distributions, but it is quite possible that the community should get stuck at a suboptimal distribution, particularly at the extreme $\langle N, 0 \rangle$. Intuitively, if p_2 is too small or N too large, there may be no benefit in a maverick's abandoning method I. The good news is that there are some instances of this general type in which the Hobbesian community not only does better than its high-minded cousins, but actually achieves a stable optimal division of cognitive labor. The bad news is that, when the community is too big, self-interest leads the community to the same suboptimal state as individual rationality.

There is a remedy, however. The trouble with large communities (more exactly communities for which kN is too big) is that a single deserter from method I cannot contribute enough effort to method II to make that method profitable. What is needed is for several people to jump ship together. Imagine, then, that the community is divided into fiefdoms (laboratories) and that, when the local chief (the lab director) decides to switch, the local peasantry (the graduate students) move, too. Suppose that each lab contains q members and that the director can thus bring it about that x members of the community switch where x is less than or equal to q. Of course, if $q > kN$, then a single laboratory can realize the intrinsic prospects of method II, and it is easy to see that there are conditions under which a stable CO-distribution is attainable. Moral: a certain amount of local autocracy—lab directors who can control the allegiances of a number of workers—can enable the community to be more flexible than it would be otherwise.

I have been exploring some of the consequences of a very schematic description of the ways in which personal motives (and social incentives) might operate in a scientific community. Obviously, my account would be improved by making more realistic assumptions and introducing factors that have so far been omitted (for example,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the possibility that both methods might succeed, individual differences in talent and interest, and so forth). I want to close this section, however, by taking up an obvious objection. Do my conclusions about the expected distributions achieved by the community not depend not only on the assumption that each of the members is driven by the desire to win the prize, but also on the supposition that each has enough information about probabilities to see how best to achieve that end? And is that supposition not highly unrealistic?

I reply that the supposition that scientists identify the probabilities that methods will succeed, given an assignment of a number of workers, is an idealization of the same kind as that traditionally made in confirmation theory. Just as we idealize the everyday judgments in which scientists assess the chances of competing theories, so, too, we can start from the ordinary sense that a method is overrepresented (or underrepresented) or from the awareness that the chances that a method will succeed are quite low unless it is pursued by a critical mass of people, and suppose that rough-and-ready judgments are replaced by the assignment of numerical probabilities.

Of course, appealing to human ambition is only the beginning of the story. Other psychological mechanisms might bring scientists closer to the CO-distribution than they would otherwise have been. Not only may vices from greed to fraud play a constructive role, but community ends may be furthered by more salubrious traits. Perseverance, personal investment, personal and national loyalties, and devotion to political causes may, on occasion, help to close a CO-IR discrepancy.

V

I have looked in some detail at the problem for methods. Can we achieve similar results for our original problem, the problem of theory choice?

Yes. But precise description of cases of theory choice turns out to be tricky. Crudely, the troubles stem from the existence of two sources of uncertainty: we need to take into account the probability that a theory will improve its apparent epistemic status and also the probability that, if it does so, it will be closer to the epistemic goal (e.g., truth). Nonetheless, if we are prepared to make some large idealizations, my original fable can be articulated in more detail. The text of the present section will tell a highly simplified version of the story; gestures in the direction of greater realism can be found in the footnotes.

Imagine that, at some moment in the history of some science, we have a pair of rival, incompatible theories, T_1 and T_2. Given the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

available evidence, the probability that T_i is true is q_i, and all members of the scientific community concerned with the theories recognize this. Suppose further that $q_1 + q_2 = 1$, that $q_1 > q_2$, but that q_1 is approximately equal to q_2.[11]

The community goal is to arrive at universal acceptance of the true theory, to eliminate the problems that currently beset this theory, and to develop the theory both in its applications to theoretical problems and to practical matters. In pursuing this goal, the community can follow one of two generic strategies: (A) assign all the scientists to T_i; (B) assign n scientists to T_1, $N - n$ to T_2 (where $0 < n < N$). I shall consider the merits of these strategies from the perspective of a later stage in the history of the community—"the time of reckoning"—at which we assign epistemic utilities to various consequences.[12]

The possible outcomes at the time of reckoning are as follows. If everyone has been pursuing the true theory, then, I assume, that is the best of all possibilities, and, through its resolution of problems, and so forth, the community has amassed epistemic utility u_1. On the other hand, if everyone has been working on the false theory, then, I shall suppose, the effort has been completely wasted, for epistemic utility $-u_1$. To understand the consequences of dividing the labor, we need to introduce the concept of a *conclusive state,* a situation in which the present standoff between T_1 and T_2 is resolved. I shall imagine that both the available theories are currently beset by anomalies, problems that it is necessary for them to overcome if they are to win unqualified acceptance (honoring the traditional idea that theories are born refuted). We would reach a conclusive state in favor of one of the theories, say T_1, just in case, at the time of reckoning, T_1 had managed to overcome its problems and T_2 had not, despite being given an opportunity to do so. The controversy between T_1 and T_2 is resolved if *both* theories are pursued and if *one* overcomes its current anomalies, while the other does not.

Now let us make the very optimistic assumption that nature, though not forthcoming, is also not hostile: although correct

[11] There are many different idealizations here: I assume that theories can be associated with definite probabilities on the basis of the available evidence, that there is universal recognition of the right probabilities, and that one of the theories is correct. One first move toward greater realism would be to relax this last supposition, allowing that $q_1 + q_2 = r < 1$.

[12] Here it would be more realistic to allow for the possibility that the community aims to use the presently available theories in achieving a more adequate descendant theory that would be closer to the truth than either of those now available.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

theories may encounter anomalies, theories that successfully overcome all their anomalies are correct. Thus, if we reach a conclusive state, then there is no need to worry about false positives—in a conclusive state we resolve the issue and we resolve it correctly. In light of this assumption, I shall assign epistemic utilities to the outcomes as follows: if division of labor by following one of the (B) strategies leads to a conclusive state, then we attain epistemic utility u_2 $(0 < u_2 < u_1)$; if it does not, then the epistemic utility is 0 (we are still in the same predicament, although our labor may have given us a clearer view of the problems that each of the rivals faces).[13]

The expected utility of the more promising (A) strategy (assign all scientists to T_1) is easily computed. It is

$$q_1 u_1 - q_2 u_1$$

To work out the expected utility of the strategy (B_n) (choose the distribution $\langle n, N-n \rangle$), we need to recognize that a conclusive state in favor of T_1 will be attained just in case (a) T_1 overcomes its current anomalies, (b) enough workers are assigned to T_2 to give T_2 a chance. I shall assume that the probability that (b) is the case is 1 if $N - n$ is larger than some value, m, and 0 otherwise. The probability that (a) obtains is the probability that T_1 is true multiplied by the probability that T_1 responds to the efforts of n scientists. Letting $p_i^*(n)$ be the probability that a true theory T_i responds to the assignment of n workers by overcoming its problems, we can write the expected utility of (B_n), where $m < n < N - m$ (so that both theories are given a chance) as

$$q_1 p_1^*(n) u_2 + q_2 p_2^*(N-n) u_2$$

Division of labor is thus preferable if there is an n, meeting the "give both a chance" constraint, $m < n < N - m$, such that

$$q_1 p_1^*(n) u_2 + q_2 p_2^*(N-n) u_2 > q_1 u_1 - q_2 u_1$$

[13] My claims about the u_i can easily be adjusted to reflect differences in views about the values of particular outcomes, or even differences in specific situations which require assignment of different values. Much more tricky is the task of replacing the hypotheses about conclusive states with more realistic assumptions. In principle, one ought to allow for the possibility that ingenuity can make a false theory continue to appear plausible, and the idea of simple opposition between victory for one theory or an inconclusive dispute should give way to study of the evolution of the probabilities assigned to the rival theories. So, in effect, we want an estimate of the probability that each theory will be assigned if there is a particular division of cognitive labor, and to use this to specify the probability of making a correct decision at the time of reckoning. My preliminary explorations of the more complex algebra that results suggest that, in a significant number of instances, the much simpler approach of the text will not lead us very far astray.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

I shall simply consider one of the forms of return function introduced in section III (similar results are easily generated in the other case). Suppose that $p_i^*(n) = p_i(3n^2 - 2n^3/kN)/k^2N^2$ $(n < kN)$, $p_i^*(n) = p_i$, otherwise. We can simplify the discussion without loss by supposing that $p_1 = p_2 = p$. Suppose, as with our earlier discussions, that $k < ½$, so that it would be possible to assign scientists to each theory in a way that would give each a maximal run for its money. If n lies in the interval $[kN, (1 - k)N]$, then it seems reasonable to conclude that the condition that both theories have been given a chance is satisfied. So the crucial inequality for the preferability of one of the (B_n) is

$$pu_2 > (q_1 - q_2)u_1$$

Thus, unless the maximal chances of a true theory overcoming its problems are low (p is small) or the utility of immediate action is high (u_1 is large relative to u_2), the CO-distribution again involves a genuine division.

Could nonepistemic incentives operate to bring individual scientists close to the CO-distribution? Let us suppose that the important motive is each scientist's desire to be singled out by posterity as an early champion of the accepted theory. If the community is initially divided, with distribution $\langle n, N - n \rangle$, and if $kN < n < (1 - k)N$, then there is a stable attainable distribution, $\langle n^*, N - n^* \rangle$, where

$$n^* = q_1N/(q_1 + q_2)$$

Provided only that $k < q_2$ (a very weak assumption, given that $k < ½$ and q_2 is close to q_1 $[=1 - q_2]$), this will be a CO-distribution. Moral: as in the case of the earlier models, there are specifiable circumstances, albeit highly idealized, in which the IR-distribution diverges from the CO-distribution and in which extra-epistemic incentives bring the community to the CO-distribution. Social structures within the scientific community can work to the advantage of the community epistemic projects by exploiting the personal motives of individuals.

VI

I want to close with some brief indications of how the analysis I have begun here might be deepened and extended. First, it would be relatively easy to consider social structures and personal motivations (national or personal loyalties, for example) that I have left out of account. Second, in my treatment of particular cases, I tacitly assumed that the size of the available workforce was fixed. But this is surely unreasonable. Just as we can think of science as facing intra-

110159

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

field optimization problems, we can also envisage issues of interfield optimization. Although personal interests may also help the community of scientists to achieve a reasonable distribution of scientists across fields, difficulties of retraining may interfere with the process.

Interfield optimization is not the end of the matter. There is a problem of division of labor of even broader scope. The epistemic goals of the community do not exhaust the set of community ends, and we can ask how, given all the aims that we have for ourselves and our fellows, we should allocate resources to the pursuit of our community epistemic goals. Given the solution to this optimization problem, we know the size of the workforce that the sciences can command. We can then ask for the optimal division of labor among scientific fields, and, finally, proceed to the question that has been addressed in a preliminary way in this essay: what is the optimal division of labor within a scientific field, and in what ways do personal epistemic and nonepistemic interests lead us toward or away from it? That question ultimately finds its place in a nested set of optimization problems.

Optimality analysis need not breed optimism. One of my main themes has been the possibility that psychological factors (and scientific institutions that exploit those factors) often thought to be detrimental to cognitive progress might turn out to play a constructive role. But it would be highly surprising if the existing social structures of science, which have evolved from the proposals of people who had quite different aims for the enterprise and who practiced it in a very different social milieu, were to be vindicated by an optimality analysis. How do we best design social institutions for the advancement of learning? The philosophers have ignored the social structure of science. The point, however, is to change it.

PHILIP KITCHER

University of California/San Diego

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

AESTHETIC QUALITIES AND AESTHETIC VALUE

TO say that an object is beautiful or ugly is seemingly to refer to a property of the object. But it is also to express a positive or negative response to it, a set of aesthetic values, and to suggest that others ought to respond in the same way. Such judgments are descriptive, expressive, and normative or prescriptive at once. These multiple features are captured well by Humean accounts that analyze the judgments as ascribing relational properties. To say that an object is beautiful is to say, in part, that it is such as to elicit a response expressing pleasure in certain observers. The observers in question must not be ignorant, biased, insensitive, or of poor taste, and they must not base their evaluations on aesthetically irrelevant properties of the objects they judge. The reference to the object's "being such . . ." captures the objective side of the relation; reference to the pleasurable response captures the expressive function of these judgments; and the ideal properties of the observers suggest that others ought to judge in the same way.

Beauty is a relatively nonspecific or broadly evaluative relational property, in that its ascription leaves unspecified how the object is such as to elicit this positive response in suitable observers. This, together with the requirement that critics not base their judgments on aesthetically irrelevant properties, implies that beauty must supervene on other properties. In general, if evaluative properties are to be analyzed in this way which captures the various functions of evaluative judgments, then they must be supervenient properties.

Before proceeding to a discussion of the base properties on which beauty supervenes, I should note that it may not be the most broadly evaluative aesthetic quality, although it is the quality most often mentioned in this light. The beautiful, as we know from Kant and other aestheticians of that age, may be contrasted with the sublime, but also with sets of artworks having other qualities that may confer artistic merit on them. Artistic merit itself is both broader and narrower than beauty: broader, in that it may be based on other properties, such as expressiveness (e.g., power) or originality, that may not confer beauty on their objects; and narrower, in that it is possessed only by artworks, whereas natural phenomena may be beautiful as well. The beautiful is pleasurable to observe; artistic merit may not always give pleasure, although it will elicit a positive response and presumably attract continued attention. Artistic merit may require study and understanding to be appreciated; beauty may seem more immediately accessible, although this is not always the case. It may

0022-362X/90/8701/23–37 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

take some musical education before one can appreciate the beauty of a Brahms symphony or a Mahler song.

Those qualities on which beauty and artistic merit most immediately supervene are themselves aesthetic and, at least in part, evaluative. I have in mind such qualities as being graceful, powerful, balanced, original, and so on. Such aesthetic qualities have been defined, for example by Sibley, as those whose apprehension requires taste.[1] This strikes me not so much as a helpful definition (to be helpful, we would need to know the nature of taste) as an indication of the evaluative aspect of these properties, again best analyzed as relations. To take taste to be a special faculty that apprehends special qualities is no more plausible than was G. E. Moore's intuitionism in ethics. Since there are no faculties or organs of taste, appeal to it signals only that ascription of these aesthetic qualities express particular tastes or sets of aesthetic values.

Appeal to these qualities supports judgments of beauty or artistic merit in a principled way, in that the presence of the properties mentioned always counts at least prima facie positively in evaluating works of art. To say that a work is powerful is to suggest that it is artistically a good work, just as to say that a man is courageous or kind, is to suggest that he is morally a good person. But this is because such terms themselves express approval. The same work that is powerful to one critic may be strident to another; a work that is intense to one may be garish to another. Aesthetic qualities are like moral properties in this regard: the man who is courageous in the eyes of one observer may be cold-blooded to another. In all these cases, the opposing evaluative properties and judgments referring to them may rest on the same nonevaluative bases. The objective sides of these pairs of relational properties are the same, but they elicit different responses expressed by the opposing terms.

A principle of beauty (or artistic merit) states that an object with a certain property P is prima facie beautiful (artistically good). This allows that P may be overridden by negative factors in the artwork, but requires that it always count in the same (positive) direction. Principles that fill in reference to the sort of properties mentioned above are not so interesting, since their existence once more may signal only that the properties in question are themselves (positively) evaluative. A far more interesting sort of principle would link evaluative to nonevaluative properties in this way. It would say that an object with nonevaluative property Q (an objective property rather

[1] Frank N. Sibley, "A Contemporary Theory of Aesthetic Qualities: Aesthetic Concepts," *The Philosophical Review*, LXVIII (1959): 421–450, p. 421.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

than a relation to a positive response) is prima facie beautiful or artistically good.

It might be objected that this distinction could be used to beg the question against interesting aesthetic principles, since any property linked always positively to evaluation could be interpreted as itself evaluative and therefore a component of only an uninteresting principle. But there is an independent way of drawing the distinction that does not beg the question. There is broad agreement on the presence of objective or nonevaluative properties that ultimately ground evaluations, and disagreement can be settled by straightforward empirical investigation. By contrast, we have seen that opposing evaluative properties and judgments expressing them may rest on the same nonevaluative bases.

If aesthetic qualities such as power, grace, and balance are (partly) evaluative, then they can once more be viewed as relations, in terms of objects being such as to elicit positive responses from suitable critics. But, unlike in the case of beauty or artistic merit, the objective sides of these relations are now more specifically indicated. These objective base properties, on which aesthetic evaluations ultimately supervene, are themselves most often relations among simpler objective elements.

Relations of different sorts form the objective bases for different evaluative aesthetic properties. 'Graceful' and 'harmonious', for example, refer to formal relations that elicit positive responses. A graceful sculpture ordinarily has smooth and flowing lines without protruding, sharply defined parts. 'Powerful' and 'soaring' refer with approval to expressive properties of the works themselves, which in turn depend on certain formal properties, although not always the same ones. A powerful piece of music, for example, cannot be legato and pianissimo throughout. 'Innovative' and 'daring' refer, again while expressing approval or positive evaluation, to relations of the work or its features to features of other works in a tradition.

The middle-level evaluative aesthetic qualities that I have been describing provide reasons for judgments of artistic merit or beauty. But ascriptions of the former properties can be challenged as well, and their defense requires further reasons that appeal ultimately to nonevaluative properties.[2] We may, if we like, follow Monroe Beardsley in defining aesthetic qualities as those which provide reasons for aesthetic evaluations (*ibid.*, p. 103); but it is most important

[2] Cf. Monroe C. Beardsley, "What Is an Aesthetic Quality?", in *The Aesthetic Point of View* (Ithaca: Cornell, 1982), pp. 104–105.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

to distinguish middle-level from ultimate reasons, and thereby relational, evaluative aesthetic qualities from nonevaluative aesthetic qualities, which we may now call basic.

This fact/value distinction is (I hope) not naive, at least in so far as one thrust of the previous discussion is that many terms of aesthetic discourse, and indeed many relational aesthetic properties, cut across it. All the judgments or ascriptions of qualities so far mentioned are both evaluative and descriptive. The fact that evaluative properties are always supervenient on others, however, implies that there must be some nonevaluative properties on which they ultimately supervene, and hence some fact/value distinction that can be recognized. Its recognition generates a constraint on rational evaluators—that the presence of all the same nonevaluative properties on different occasions must elicit the same aesthetic evaluations. But this minimal constraint leaves it open whether rational evaluators can disagree about the same works (indeed, the notion of taste indicates that they can), and whether there are any nonevaluative properties that always count in the same way toward evaluations, that is, any interesting principles of taste.

I have agreed with Beardsley that aesthetic qualities are those which provide reasons for aesthetic evaluations, but I have pointed out that many of these properties are themselves evaluative, relational properties that must supervene ultimately on nonevaluative, basic aesthetic qualities. We can characterize basic aesthetic qualities in ways that do not explicitly refer to the category of the aesthetic. Such qualities are first of all phenomenal properties, those which appear in perceptual experiences, and relations among phenomenal properties. Furthermore, at least in the case of aesthetic properties that ground positive evaluations (those on which I shall concentrate below), the experiences in which they appear are thereby made more valuable. In saying this, it is important to emphasize that the value that attaches to experiences of aesthetic qualities cannot be detached from the qualities at which they are directed. Aesthetic value does not lie in the mere pleasantness of the sensations created when enjoying a work of art, even when there are pleasant sensations involved in such enjoyment, as in listening to a beautiful piece of music. The perceptual experiences of good artworks are valuable because of the way they are structured when directed at aesthetic features of those objects. Works with artistic merit are such as to elicit positive responses in virtue of the way that their phenomenal properties and relations among these properties generate experiences of them.

Several objections may be raised to the claim that aesthetic quali-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ties are phenomenal properties. One derives from a phenomenon emphasized in the recent literature, most emphatically by Arthur C. Danto,[3] that of two perceptually indistinguishable objects that merit different aesthetic descriptions and evaluations, even from the same critics. An example of such a pair would be an ordinary urinal versus Duchamp's perceptually indistinguishable work of art. If perceptually indistinguishable objects can be such that one is a work of art (and of high artistic merit according to at least some critics) and another is not, then artistic merit cannot be grounded entirely in perceptible, phenomenal properties. Indeed, Danto's examples may seem to cast doubt also on the minimal constraint on aesthetic evaluation proposed above: that repetition of all the same nonevaluative properties must elicit the same evaluations from the same rational critics. This doubt will be dispelled below as well by defending the analysis of aesthetic qualities as phenomenal properties and relations among them.

A second objection to that analysis notes that representational and expressive functions of artworks enter into their artistic merit, and that once more these are not simple phenomenal properties of the works in themselves. To grasp fully the representational and expressive features of certain works may require knowledge of human nature and history far beyond anything directly presented in the works themselves. In order to appreciate Tchaikovsky's *1812 Overture* or Picasso's *Guernica,* one must know something of the glories and horrors of war, and something of the histories of the particular wars depicted. In novels, as opposed to visual art and music, it might plausibly be claimed that phenomenal properties are of minimal significance, as opposed to the semantic properties of the language by means of which it represents and expresses. The structure of a novel is defined not by relations among phenomenal elements, such as tones or colored patches, but rather by elements of plot and character depiction as these develop and interrelate. In some novels, these relations may be quite subtle, determined, for example, by shifting points of view among narrators and other characters. Thus, even an aesthetician who takes self-contained formal properties of artworks to be of paramount aesthetic value cannot view these properties as always phenomenal.

What these objections fail to note is the claim that aesthetic qualities consist not only in phenomenal properties, but also in relations among them, relations that are not always perceivable. In the remainder of this paper, I shall attempt to clarify the nature of the

[3] *The Transfiguration of the Commonplace* (Cambridge: Harvard, 1981).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

relations that are aesthetically significant, and then suggest why properties that constitute them should be a source of aesthetic value. Finally, I shall comment on whether the connections between these relations and the positive value judgments they elicit can be captured by a set of aesthetic principles.

I

The first set of aesthetically significant relations hold among phenomenal elements within works themselves and define their internal structures. (At present I limit discussion to visual arts and music; the discussion of literature will be continued below.) The elements here consist in tones in musical pieces, colored patches or lines in paintings, surfaces or chunks of sculptures, and so on. The relations include contrasts, variations, repetitions, similarities, blendings, attractions, repulsions, tensions, and resolutions, and developments based on various of these other relations. They create structures referred to as dynamic, tightly knit, balanced, harmonious, and so on. These structural features are perceivable, and they hold among elements as phenomenally, not physically, defined. That the tonal progression from the tonic to the dominant and back to the tonic chord constitutes a development in terms of tension and its resolution is not inherent in the physics of the vibrations, but in how the tones appear to us and cause us to respond perceptually and emotionally. Such formal features can be a source of aesthetic value in themselves, apart from any representational functions of the works that contain them, and even apart from whatever emotions might be expressed by means of them. Although some formal features in themselves are expressive, not all that are of aesthetic value need be. The grace, balance, and harmony of an abstract sculpture, for example a Gabo, can have aesthetic value without expressing anything to the viewer.

Nevertheless, it is undeniable that a second source of aesthetic value lies in the representational and/or expressive character of some artworks and their properties. Here we must recognize relations of a work's elements and formal properties to perceptual and emotional experiences outside the work. The manner in which an artwork represents may be quite straightforward, especially in descriptive literature and visual art, although even in the latter case there will be representational conventions at work to supplement resemblances. Such conventions may go unnoticed by audiences trained to interpret them, and sometimes even by artists who use them. Representation in music, beyond such devices as imitating bird calls or other sounds, is more complex, always shading into expression, which is generally subtler in its methods.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

I mentioned above terms that refer to evaluative, expressive properties, such as 'powerful' or 'soaring'. Such properties, I suggested, can be analyzed in terms of their objects' being such as to elicit positive responses in virtue of certain other expressive properties. We also find, of course, terms that refer nonevaluatively to expressive properties themselves, such as 'sad', 'tense', 'somber', 'cheerful'. There is no sharp dividing line between these types of terms, and context may determine whether approval or disapproval is being expressed by the latter. It is clear, however, that, whichever of these terms we are using, we refer, at least in part, to expressive qualities in the works themselves. The expressive qualities referred to relate to human emotional states, but not in the simple sense that they express emotions felt by the artist in creating the work or by the audience in perceiving it. Instead, the audience is to respond to the work in a way similar to that in which it responds to the emotion when encountered elsewhere.[4] The proper response to an emotion is not always simply to feel the same emotion: one responds to sadness, for example, not simply by feeling sad, but also with sympathy and pity.

The way one responds to emotion in art is only similar to one's responses in real life, in that one is to react to artistic expression without believing that the emotion expressed is literally present, without acting in a way normally appropriate to such a response, and without losing awareness of the elements and formal properties of the work which somehow express the emotion in question. How this effect is achieved differs from genre to genre and from work to work. In straightforwardly representational art, the method may be equally straightforward. Representing or describing a sad person, for example, may capture the person's sadness and so elicit the appropriate response. In music and abstract art, the relations are again more complex. Emotions may be expressed through representation, in the formal structure of a work, of the structural properties of typical expressions of the emotions. Sad music, for example, typically has slow, flowing lines; cheerful music, more rapid rhythms; whereas powerful pieces build tension through tonal development and volume. It is more mysterious why minor chords or dark, muted colors of paintings in themselves appear sad. There seems to be no single method by which music, let alone other arts, express.

The relations that constitute expression and representation in art are not themselves perceivable. But I think it fair to say that they

[4] Cf. Roger Scruton, *Art and Imagination* (London: Methuen, 1974), p. 128.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

relate phenomenal, or at least experienced, elements. We respond to sadness in music not because we believe that the music is (literally) sad, but because it recalls to us the experience or observation of another's sadness or its expression. As mentioned, the other side of the relation, the properties in the music which express the emotion, must remain in focus as well. We must respond to these phenomenal properties as expressive of sadness. If we are to react properly, to sad representational works as well, we must remain aware that the sadness is not real, but only expressed in the experienced properties of the works. As Danto notes (*op. cit.*, pp. 147–148), when art is representational (or expressive), the manner of representation (or expression) always matters aesthetically. And the manner or method always involves phenomenal elements and formal properties within the works. Finally, the emotions expressed or objects represented in artworks are expressed or represented as the artist has experienced them. We are to experience them and respond through the eyes of the artist, so that our way of experiencing or perceiving may be altered or expanded.

The third set of aesthetically significant relations consists in relations of properties in artworks to earlier artworks that define a tradition. These relations are referred to by terms such as 'original', 'innovative', 'conservative', 'daring', and so on. They indicate a historical place for an artwork within a developing sequence. Works may be aesthetically valuable solely because of the way in which they continue, modify, overthrow, or extend a particular tradition within a particular genre. Once more we can consider such relations to hold between the phenomenal properties of a present work and those of earlier (and perhaps even later) works. Once more, appreciation of the relevant historical relations informs and affects the way we perceive phenomenal properties within the work, upon which attention must always remain focused. This effect may be for better or worse. Present audiences probably cannot feel the power of a Haydn finale as audiences in Haydn's time did, the former having absorbed more thoroughly the symphonies of Beethoven. On the other hand, we hear the development section in the first movement of the *Eroica Symphony* differently when we recognize that new ground was being broken in the history of music.

This last sort of relation holds between phenomenal properties of artworks perceived at different times, so that once again the relations themselves are not perceivable. The properties related, however, are phenomenal properties of the works themselves, for example formal properties, as well as expressive properties that I have defined also as complex relations.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

II

Having included relations that may be themselves unperceivable in the analysis of aesthetic properties, we are now ready to answer the objections raised earlier. Reconsider first Danto's examples of perceptually indistinguishable objects that are aesthetically different. A natural object perceptually indistinguishable from an artwork, say a sculpture, is not itself a work of art. But then it does not bear those relations which we have recognized to be aesthetically significant. It does not express or represent anything, for example, although it may have formally pleasing properties and so be beautiful (*op. cit.*, pp. 93–94). A copy that is perceptually identical to an original work of art also differs in its aesthetic relations. It does not express or represent through the eyes or experience of its creator. Nor does it ever bear the same relation to the tradition. Whereas the original may continue or modify the development of the genre in an interesting way, the copy does not continue or extend the tradition at all (once we include the original in that tradition).

Other objections appealed to literature, where the elements that enter into formal, expressive, and historical relations seem to be defined according to the semantic content of the writing, rather than phenomenally. Furthermore, there may be value in what a novel says—the moral, political, or philosophical truths contained therein. Nevertheless, although content certainly serves to define character, plot development, and hence structure in a novel, the structure, characters, and even moral and philosophical content of a novel have aesthetic value only when they serve to inform our experience in reading it. The aesthetic qualities of fiction do not lie in the historical or philosophical truths that it might convey. A novel makes great literature not if it describes a true moral theory or historical era, but if the moral beliefs and experiences of the characters, or the historical circumstances in which they find themselves, help to define them and relations between them in an aesthetically fulfilling way.

Even the formal structure of the novel must structure our experience of it, must help us to understand the characters and to experience through their points of view. Just as a bit of philosophical or historical exposition cannot substitute for the functions of these elements in a novel, so a diagram or chart, which may represent the structure in a novel with perfect accuracy, cannot substitute for the way the structure affects our experience in reading it. All these points supplement appeal to the sensuous nature of the language itself in a good piece of fiction or poetry, which once more may be of aesthetic value in the quasi-musical forms it creates and in the way it may intensify our experience in reading the work.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The analysis of aesthetic qualities as phenomenal properties and certain relations among them stands against the objections raised earlier. It remains first to inquire why and how these relations constitute sources of aesthetic value. Although I do not deny that isolated elements of artworks or natural objects, for example certain colors or tones, may be so pleasing to the senses as to merit the ascription of beauty, certainly the greatest sources of aesthetic value lie in those relations described above. And whereas the question of why certain colors or tones are found to be so pleasant is one for psychologists to answer (if, indeed, there is an answer here, as opposed to a primitive psychological fact), it is of more philosophical interest to explain how aesthetic relations create aesthetic value, to suggest (for further development) a theory of artistic value.

The first thing to note here is the interaction among elements within these relations and the interaction among the relations themselves. Regarding the former, elements of artworks are transformed when relating to others in broader structures. We hear a tone, for example, completely differently when it is part of a chord or melody, and a chord differently depending on the key and modulations in which it is embedded. And it is well-known that we perceive colors differently depending on the surrounding colors. This is one, but only one, reason why it is doubtful that we can find nonevaluative properties that always contribute in the same way to aesthetic value. If simpler properties are altered by the relations into which they enter within artworks, then they will not always maintain the same values. Only broader relations themselves, then, can be candidates for entering interesting aesthetic principles. But the question is whether we can find anything short of whole artworks that must be judged in the same way on different occasions of appearance, any principle more interesting than the minimal constraint on rational evaluation suggested in section I.

I have indicated in passing the many ways in which the relations deemed aesthetically significant interact with each other. We even have terms of positive evaluation for such interaction, such as 'aptness of form to content'. Form determines representation and expression, and any of the three may determine the historical place of a work within a tradition. In music, expressiveness may be achieved through formal qualities of rhythm, pitch, and tonal development; in painting, representation may elicit emotional responses similar to those made in reaction to the objects represented. But these interactions are not always one-way. I noted that the historical place of a work, for example a Haydn symphony, may influence for better or worse its expressive power for a contemporary audience. Further-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

more, expressive qualities can help to determine form; modulation takes the form of felt tensions and resolutions in many tonal musical pieces. The variations here seem almost endless and can acquire such complexity and subtlety as to become the controversial topics of critical interpretation. Satirical literature, for example, can represent nonsatirical counterparts in the literary tradition, whereas some contemporary paintings represent their own places within recent movements in painting. Here self-conscious historical relations help to determine representation, what the work is about.

Before commenting on how these interactions create values in artworks, we may note why the three sorts of relations in themselves should be sources of aesthetic value. Form or structure constitutes order among distinct elements which allows intellect or perception to grasp them in more significant chunks. Since order is sought by both faculties, we should not be surprised that it is pleasing when found, more so when found after a challenge. As other aestheticians have commented, order within complexity is therefore of particular value. Relations of an artwork to its tradition represent a different kind of order that allows for greater comprehension of the work as a whole.

As form allows for the exercise of intellect and perceptual comprehension, so representation and expression in art allow for the exercise of imagination and emotional capacities without the personal costs often associated with the latter. Once more, such exercise, or "free play," as some aestheticians call it, both pleases in itself and has instrumental value in helping to develop the faculties or, in more contemporary terms, the capacities in question. Through exposure to art, we can become more perceptually aware, more open to various sensuous enjoyments, and more emotionally sensitive and sensible as well.

The greater value of these relations lies in the way that they affect perceptions of the elements within artworks. I have emphasized that form, representation, expressiveness, and even originality must impregnate our experience of the sensible properties of artworks in order to be of aesthetic value. We hear notes and chords as elements within melodic and harmonic structures, and they have meaning for us in those terms. These structures are expressive to us, and the emotions they express also imbue the hearing of particular chords with a deeper significance. We must hear the sadness in the music, for example, rather than simply be caused to dwell on sadness, which would have negative, if any, value. The range of emotions expressed in a single symphony may condense those felt over long periods of nonaesthetic experience. This intensity is even more evident in read-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ing literature, when we can vicariously experience lifetimes in a matter of hours.

Experience of artistic elements which is imbued in these ways by emotion or structure or historical significance is thereby made richer, more meaningful, intense, and condensed. When interactions among these different relations inform our perceptions of the elements within artworks, the effects are yet more remarkable, and the value to be derived is itself more intense. The value of art, on this view, lies largely in the richness of the perceptual and emotional experiences it affords us, and the theory of aesthetic value explains the sources of this richness. I have only indicated the outlines of such explanations here, but we must proceed at this point to our final question.

III

If experience richly imbued with meaning through structure, expression, and historical significance is valuable to us, why should we not be able to capture these sources of value in a set of aesthetic principles of the sort indicated above (section I)? Interesting principles, we noted, must link nonevaluative to evaluative aesthetic properties; but if structural, expressive, and historical relations can be described in nonevaluative terms, why should there be any great difficulty in specifying these links? If these relations create value in the experience of elements within artworks, why should they not do so in lawlike ways?

There are at least two reasons, both alluded to briefly above. First, elements are transformed by the relations into which they enter, and so do not themselves always count in the same (lawlike) way toward the value of works in which they are found. Entire works in all their historical relations must be judged alike on different occasions of apprehension, but this constraint fails to generate any aesthetic principles, since entire works in all their relations are never repeated. Given the transformation of elements, it is difficult to imagine why any part of a work short of the whole could not be altered in its value when placed in a different context. Nevertheless, one might maintain, although elements may be transformed by relations into which they enter, they can still be transformed in lawlike ways. If certain relations add value to the elements in artworks that they connect, why can these value creating transformations themselves not be lawlike, hence expressed in aesthetic principles? The answer here lies in irreconcilable differences in taste.

Some aesthetic disagreements result from inattention by one of the parties to certain phenomenal properties in the works themselves. These disputes can be resolved simply by calling attention to

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the properties in question. Other disputes rest on failure to appreciate certain of the value-creating relations that may not be themselves perceivable. Appreciation may require knowledge of relevant forms, expressive or representational conventions, or traditions, and some disagreements can be resolved, at least in the long run, by education of one of the disputants in the relevant relations. There will remain, however, a residue of disagreement that is not resolvable in these ways. The relevant relations may inform the experiences of different evaluators of a work in different ways, and hence the responses elicited may differ as well, even after exposure to the work and education. There is room for differences among educated tastes. More frequently, different evaluators place emphasis on different relations at the expense of others. This may result in distinct interpretations or simply in opposed evaluations. For some music critics, for example, the formal and expressive properties of pieces by Saint-Saens give them great aesthetic value; while for others, his failure to extend or alter the course of the Romantic tradition in an interesting way robs his music of aesthetic value and renders it insignificant.

These two features of aesthetic qualities—that phenomenal properties are transformed by relations and that the latter affect different evaluators in different ways—defeat attempts to specify interesting aesthetic principles. The two features themselves are connected, in that elements take on different significance depending on the emphasis placed on particular relations they may enter. The result is that evaluative aesthetic properties seem to supervene on nonevaluative properties without being necessitated by them. To say that they supervene is to repeat the constraint on rational evaluation that we have noted: there can be no change in a rational critic's evaluation of a work without some change in its nonevaluative properties. To say that evaluative properties are not necessitated by nonevaluative ones is to allow for differences in evaluations of the same work by different critics, to allow for disagreement without error or even insensitivity.[5]

When aestheticians argue in favor of aesthetic principles, they often have in mind support for broad judgment of aesthetic merit or

[5] Cannot a single rational evaluator change taste and thereby violate the constraint to judge the same work in the same way on different occasions? Not, I think, without holding her former judgment to be mistaken, to have missed some source of value in the work or to have mistaken for a source of value what on more prolonged inspection produces only tedium. But then, must she not hold opposing judgments by others to be equally mistaken, and so fail to allow for disagreement without error? I believe that one can recognize irreconcilable differences in taste while regarding each change in one's own taste as an improvement.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

beauty by appeal to what I have called middle-level, evaluative, aesthetic properties, such as grace or power.[6] When they argue against principles, they may have in mind the absence of necessary connections between basic, nonevaluative, aesthetic qualities and rational evaluation, which is evidenced by irreconcilable disagreements among knowledgeable and sensitive critics. Parties to such disputes in aesthetics obviously can talk past each other. I have granted the existence of the less interesting sort of aesthetic principle, while denying the more interesting kind, and explaining both features of aesthetic reasons.

The impossibility of specifying interesting aesthetic principles, either narrow or broad, is well-illustrated by recent attempts of aestheticians to do so. We may consider first a single example from David Pole.[7] He proposes as a case of a nonevaluative description that entails a (prima facie) evaluation, a character without apparent motive inserted into a play for the sole purpose of moving the plot along (*ibid.*, p. 154). In such contexts, the lack of motive is a defect, Pole maintains, and necessarily so. He recognizes, of course, that this feature counts in only a prima facie way toward a negative evaluation of the play, as is obvious from the fact that his example is from a well-known play by Shakespeare. But the problem is instead that the generalization is not perfect. In a play such as *Waiting for Godot*, lack of motive is a virtue, at least in so far as it is part of the point. This is not to say that all critics must approve of that play, but, insofar as they do, they will be unlikely to disapprove of the lack of apparent motives in the characters. Nor can we find a non-question-begging relevant difference between the plays which accounts for the different effects of this feature and which we can generalize. All we can say is that *Godot* is not the sort of play in which lack of motives counts negatively. Clearly that sort of statement will not do as a way to save the law or aesthetic principle.

The failure of Pole's principle can be explained by either or both features of aesthetic qualities to which I have drawn attention. The lack of motive that may be a defect in the context of one play is not when related to the other elements in a different play. Both of these plays, but certainly *Godot*, may elicit approval from some rational critics and disapproval from others; and such differences will transfer to this particular feature of the play as well.

The same factors defeat attempts to find broader aesthetic laws,

[6] See, for example, Sibley, "General Criteria and Reasons in Aesthetics," in *Essays on Aesthetics*, John Fisher, ed. (Philadelphia: Temple, 1983).

[7] "Art and Generality," in *Aesthetics, Form and Emotion* (New York: St. Martin's, 1983).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

too. One of the boldest recent attempts at reducing aesthetic to nonaesthetic qualities is that of Guy Sircello,[8] for whom beauty consists in having certain other properties to a high degree. Properties that admit of qualitative degrees relative to objects that have them are beautiful when possessed to a high degree, according to his theory of beauty. His paradigm is vividness in colors. Colors are beautiful in respect to vividness (but perhaps not in other respects) if they are very vivid.

The theory fails to specify a law, let alone a reduction, even in its paradigm case. On the one hand, a color that is beautifully vivid to one observer may be simply garish to another, not in the sense that its beauty is overridden or defeated by its garishness, but in that it is simply garish instead of beautiful. On the other hand, a pastel may be beautiful without our being tempted to say that it is a vivid pastel (whatever that would mean) (cf. *op. cit.*, p. 31). Then, too, the same color that is beautiful in the context of one color scheme may be offensive in the context of another. In the latter context, there is no reason to say that it remains beautiful in itself; rather, beauty of color, like perception of color, seems to be in part a function of context.

If this theory fails in its paradigm case, then there is no need to press the more obvious objection that qualities like sliminess become more rather than less distasteful the greater the degree to which they are present. Sircello attempts to answer this objection by saying that we may not be good judges of beauty in such properties (*op. cit.*, p. 70). But then who is to judge such beauty, slugs and catfish? He also hedges the general thesis by claiming that it holds only of qualities that are not defects in their objects. Once more, however, this seems to say that the law obtains except where it does not, which, of course, is to admit that it is not a law at all.

Examples could be multiplied. I use these as instances of the ways in which I believe that any attempt to state interesting aesthetic principles will fail. Those principles which do seem to work, I have maintained, refer to evaluative aesthetic qualities throughout. The absence of laws, however, does not entail that we cannot have an interesting theory of aesthetic qualities and aesthetic value. I hope that I have suggested such a theory. We can recognize the ways in which basic aesthetic qualities constitute sources of value without attempting to provide formulas for artists.

ALAN H. GOLDMAN

University of Miami

[8] *A New Theory of Beauty* (Princeton: University Press, 1975).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

BOOK REVIEWS

Thought and Reference. KENT BACH. New York: Oxford University Press, 1988. ix, 305 p. Cloth $59.00.*

Recent work in the philosophy of language has shown an enlivened interest in the distinction between the genuinely semantic features of a language and the various purely pragmatic features that attend its use. The difficulties involved in drawing this distinction are, however, not always fully appreciated. Thus, for example, it has become an increasingly popular strategy to argue that, appearances to the contrary, one or another semantic proposal is really correct—the contrary appearances being explicable by pragmatic considerations. But such attempts to exploit the semantic/pragmatic distinction often limp precisely because the distinction is drawn too crudely or in a theoretically self-serving way.

Both the distinction and this strategy play central roles in Kent Bach's *Thought and Reference.* Armed with them, he offers an ambitious and provocative account of the semantics of various referring expressions, their use to refer, and the thoughts they are used to express. The book's principal thesis is that reference is an essentially pragmatic phenomenon. To refer to something is, according to Bach, to use an expression with the intention—always part of an inclusive illocutionary intention—to refer one's audience to the object of the thought one is expressing (51). Our ability so to use these expressions, however, should not constrain, in the way traditionally assumed, the semantic accounts we give them. Bach's astonishing view is that an expression's referring function, when it has one, never makes a semantic contribution to the sentences in which it occurs (58).

Bach's book begins with an account of what he calls "singular thoughts"—thoughts we aim to express when we utter a sentence with an intention to refer. After quickly outlining the speech-act theoretic framework he employs, Bach proceeds to detail and defend his own, often surprising, semantic accounts of various expressions we standardly use to refer—definite descriptions, proper names, indexicals (in both their deictic and anaphoric uses) and even, ac-

* I am grateful to David Sachs, George Wilson, and Stephen Yablo for their suggestions.

0022-362X/90/8701/38–45 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

cording to Bach, various quantifier phrases. There is, in addition, a discussion of the role of referring expressions in belief contexts, and a final chapter on reference and natural kinds. Bach's main aim in developing these different accounts is to clarify and demonstrate the way in which reference is supposed to be essentially pragmatic.

The adequacy of Bach's semantic proposals and, ultimately, the significance of his main thesis, depends on how he draws the line between those features of language use which are due to genuinely semantic considerations and those which are primarily due to pragmatic considerations. Although Bach tells us a lot about where he wants to draw this line, he unfortunately offers little or no convincing justification for why we should draw it where he says. In consequence, most of Bach's semantic proposals are not likely to win acceptance, and the import of his claim that reference is essentially pragmatic remains quite uncertain.

The core of Bach's book is his discussion of the semantics and pragmatics of the various expressions we standardly use to refer. Paradigmatic of Bach's strategy here is his treatment of definite descriptions and proper names. Since he claims that a proper name, *N*, is semantically equivalent to the definite description, 'the bearer of "*N*" ' (135), his discussion of descriptions is especially central. Bach's view is that Bertrand Russell was right: sentences containing definite descriptions or ordinary proper names literally express general uniqueness propositions of the sort that Russell's theory of descriptions says they express. Acceptance of Russell's theory, together with Bach's "nominal description theory" of names (NDT), precludes definite descriptions and proper names from being viewed *semantically* as referring expressions. A speaker, in literally uttering a sentence containing a description or name, is not thereby obliged to (purport to) refer to anything. Their referential function on such occasions must, consequently, be explained pragmatically and not in terms of the semantic properties of these expressions.

Bach dismisses a host of apparent difficulties with and familiar alternatives to Russell's theory, arguing that what appear to be exceptions to or counterintuitive consequences of the theory can, by and large, be explained away or accommodated by an appeal to pragmatic considerations.[1] One of the most serious difficulties is

[1] Although he disputes certain features of Saul Kripke's proposal, Bach takes as his mode here Kripke's more circumscribed defense of Russell against Keith Donnellan. By challenging the semantic status of Donnellan's distinction between refer-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

posed by the fact that nearly all of the definite descriptions used in ordinary speech are "incomplete"—they fail to be uniquely satisfied. And since most proper names are shared by more than one bearer, descriptions of the sort Bach claims are semantically equivalent to proper names will also, typically, be "incomplete." It would seem, then, that Russell's theory has the strikingly counterintuitive consequence that nearly everything we say when we literally utter sentences containing incomplete definite descriptions or shared proper names is false. We do not, however, standardly regard these utterances as false—and certainly never just because they contain incomplete definite descriptions or shared proper names. If, however, it is a direct consequence of a semantic account of some very commonly used type of expression that most of the competent utterances of sentences containing expressions of that type turn out to be false, and if, furthermore, no reasonable explanation is available of why the speakers should so often and so consistently be in error, then we have ample reason to reject the account.

Bach, however, is not deterred (103 ff.). He accepts that most of the sentences we utter which contain incomplete definite descriptions and shared proper names are literally false. He hopes, however, to avoid the difficulties just noted by insisting that such sentences are *standardly used nonliterally*. Not only are most of the sentences we utter containing descriptions or names literally false, but in uttering them, we only rarely mean what we say! However counterintuitive this may appear, it does not, according to Bach, raise serious difficulties for Russell. In order to sustain Russell's analysis, we need only provide a plausible pragmatic account of how we might standardly mean whatever we do by using sentences that mean what Russell says. But given the admitted absurdity of supposing that we might mean what, according to Russell, a sentence with a transparently incomplete description or shared name means, such an account is, according to Bach, easy enough to provide. Indeed, it is precisely because the absurdity is so obvious that we reasonably can, and typically do,

---

ential and attributive uses of definite descriptions, Kripke argued that the mere existence of referential uses of definite descriptions poses no decisive threat to Russell's theory. Bach's hope is, in effect, to generalize both Kripke's strategy and his results. See Donnellan, "Reference and Definite Descriptions," *Philosophical Review*, LXXV (1966): 231–304; Kripke, "Speaker's Reference and Semantic Reference," *Contemporary Perspectives in the Philosophy of Language*, P. French et al., eds. (Minneapolis: Minnesota UP, 1977), pp. 6–27.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

expect our hearer also to appreciate it and thereby take us to be speaking nonliterally. And, furthermore, on the basis of that appreciation together with other contextual clues, we can also count on her being able to infer what exactly we do mean—and, in particular, what object we intend to be talking about.

If it were really as easy as this suggests to "accommodate" prima facie counterintuitive semantic theses about commonly used expressions by relegating to pragmatics the account of what speakers standardly mean in uttering sentences containing them, then it would be doubtful that the mere availability of such pragmatic accounts ought to carry much probative weight in semantic disputes. After all, what is to prevent Bach's philosophical opponents from similarly attempting pragmatically to insulate their alternative semantic theses against what he would regard as their counterintuitive consequences? Indeed, certain direct reference theorists have recently appealed to pragmatic considerations to argue that the substitution of coreferential expressions in the content clauses of propositional attitude ascriptions does, contrary to appearances, preserve truth value. The contrary appearances are to be explained by pragmatic considerations.[2] These very "contrary appearances," however, constitute, for Bach, an especially important motivation for rejecting direct reference approaches and adopting his alternative view.

Bach's readers are likely to wonder what plausible general principles could validate the choices he makes between the alternative ways one might "divide up" the relative contributions that semantic and pragmatic factors make in determining the patterns or use associated with the referring expressions he discusses. This should in turn lead them to wonder whether genuinely plausible pragmatic accounts of the relevant phenomena are really as available as Bach's discussion suggests. For although one may want to allow that, in appropriately recherché circumstances, a speaker can mean nearly anything by uttering nearly any sentence, one may also want to question just how obvious it is that—short, perhaps of assuming that special conventions are in place—speakers could *standardly* mean to be expressing singular thoughts by uttering sentences that literally express Russell-style general uniqueness propositions. But if plausibility recommends that special conventions are operative in connecting the utterance of sentences containing incomplete descriptions or shared

[2] See, for example, the discussion in Nathan Salmon, *Frege's Puzzle* (Cambridge: MIT, 1986), esp. pp. 78–79, 84–85, 114–118.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

proper names with the expression of singular thoughts, what shows that they could not or should not be semantic conventions?

If Bach is to validate the probative superiority of his proposed pragmatic accommodation over those of his actual and potential opponents, then he owes us a much more systematic and compelling account of the interrelations between the semantic and the pragmatic than he has provided. Some fairly pressing difficulties with Bach's actual proposal, however, suggest that no plausible account of the desired kind will sustain his attempted pragmatic rescue of the Russellian theory.

As nearly everything here hangs on Bach's claim that we standardly use sentences of the problematic kind nonliterally, one might expect him to spend some time motivating and defending an account of how we might independently determine when an utterance is literal or not. But here we will be disappointed.[3] It will not do to say that utterances of sentences containing incomplete descriptions or proper names must be nonliteral, given their patent falsity, for this would simply beg the question in favor of the Russellian analysis. If Russell's analysis were correct, such sentences would generally be transparently false. There is, however, no convincing independent evidence that they are, or that speakers take them to be so.

If a sentence *P* is transparently false for some speaker *S*, then, whether or not *P* is standardly used nonliterally by *S*, we ought in general to be able to expect *S*'s unhesitating assent to literal utterances of instances of the following scheme: It is not the case that literally *P*. Whereas the sentence, 'I have nothing to wear'—one of Bach's favorite examples of a sentence "standardly used non-literally" (79 ff.)—clearly passes this test, standardly used sentences containing shared proper names and incomplete descriptions will usually fail it.[4] Bach neither considers nor provides the materials for

[3] Bach does offer what he calls a "test for standardized non-literality" (81), but this test proves to be of little use in the present context. According to this test, in order to show that a standard use *M* is not a literal use of an expression *e*, we must (at least) be able pragmatically to explain use *M* on the assumption that some other use *L* is *e*'s literal use. Since it is just the availability of a plausible and probative pragmatic account of this sort which is at issue, we cannot assume, in giving it, that this test has already been satisfied.

[4] If one were to ask in response to a typical hyperbolic use of 'I have nothing to wear'—"Is it literally true that you have nothing to wear?," a negative response is to be expected. But if, alternatively, upon being told under normal circumstances, "Dinner is on the table" or "The telephone bill arrived today," one were to ask, "Is it literally true that dinner is on the table?" or "Is it literally true that the telephone

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

accommodating this fact. But unless speakers do standardly regard sentences containing incomplete descriptions and proper names as transparently false, the sort of communicative expectations, intentions, and inferences required by Bach's pragmatic account will be unavailable.

The inadequacy of Bach's proposal is further manifest by NDT's inability to pass the following test for semantic equivalence—a test, as far as I can tell, recommended by principles Bach explicitly endorses. If two sentences are semantically equivalent, then precisely the same claim will be made by the (literal) assertoric utterance of either of them under otherwise pragmatically identical circumstances. Consequently, if, upon receiving certain information, a speaker were led to doubt, deny, retract, etc., the claim he would have made uttering one of the sentences, then he should react the same way to the claim he would have made uttering the equivalent sentence under otherwise pragmatically identical circumstances. If, on the other hand, it is clear that, under those circumstances, a speaker would systematically diverge in his reactions, this can be taken as compelling evidence that the relevant sentences were not in fact semantically equivalent.

NDT fails this test, as the following example shows. For Bach, the following two sentences are, in the relevant sense, semantically equivalent:

(1) Tweety has a spot on its tail.
(2) 'Tweety' has one and only one bearer and it has a spot on its tail.

Suppose that Ralph received a canary for his birthday, and, wanting to give it a unique name, decided to call it "Tweety." Ralph now believes that his canary is the only bearer of the name 'Tweety'. Now imagine a situation in which Ralph utters (2). Assume that, given his beliefs, Ralph intends his utterance to be literal. Suppose though that, after uttering (2), Ralph learns that 'Tweety' is not a unique name after all. Under the circumstances, we would surely expect him to be willing to concede that his claim was not true. Indeed, there should be no hesitation in his now assenting to the negation of (2). Suppose, however, that, under otherwise pragmatically identical circumstances, Ralph had uttered (1) instead, and then, in similar fashion, came to learn that 'Tweety' is not a unique name? It is hardly

---

bill arrived today?," surely one would not normally expect or get a similarly negative reply.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

credible, however, to suppose that Ralph would, or should, now be willing to deny that what he had claimed, in uttering (1), was true. But if this is right, sentences containing proper names should not be supposed to have the semantics Bach assigns them.[5]

Bach's Russellian theory of descriptions and NDT face many hurdles besides those I have raised. Bach addresses a variety of these and satisfies himself that they do not successfully threaten his proposal. If, however, for the reasons I have offered, Bach's account is incapable of surmounting the fundamental difficulties raised by descriptive incompleteness, then his responses to these further hurdles will be largely immaterial.

Bach's strategy of pragmatic accommodation does not always require us to suppose that the standard uses of sentences containing the referring expressions are nonliteral. Bach does allow that the standard uses of certain sentences may very well be literal—though, in such cases, we are, for various reasons, nearly always required to distinguish between what the speaker *primarily* means—that is, what he means primarily to express by using the sentence literally—and what he means merely insofar as it is a literal use. By thus distinguishing between what the sentence literally means and what the speaker primarily means in literally using it, Bach clearly hopes that he will have provided the pragmatic space necessary to explain away the awkward-seeming consequences of his semantic proposals and secure the essentially pragmatic status of reference. If I am right, however, arguments very much like those I have used to challenge Bach's pragmatic attempt to rescue his Russellian theory of descriptions and names can be used to raise similar doubts about his other semantic proposals and their attempted pragmatic defense.

[5] A related point can be made a propos incomplete descriptions. Consider the following sentence: (1) The telephone bill came today. According to Bach, this is semantically equivalent to the merely more prolix, (2) There is one and only one telephone bill and it came today. Imagine someone uttering (1) under the sort of circumstances in which it would be normal to do so. Now imagine their uttering (2) under otherwise pragmatically identical circumstances. One would certainly react to these two utterances differently, for plainly the utterance of (2)—though not of (1)—would, under the circumstances, be patently odd. Is it plausible here to credit the differences in one's reaction merely to the prolixity of (2)? When we are struck by the oddity of someone's uttering (2), under circumstances appropriate for the utterance of (1), are we merely puzzled by why the speaker should choose so prolix a way of speaking? Reflection, I propose, suggests not. It seems, rather, that with respect to the utterance of (2), though not with respect to the utterance of (1), we are likely to be puzzled why, under the circumstances, the speaker should choose to utter a manifestly false sentence. As the circumstances are assumed to be the same, to what besides a semantic difference, are we to credit this differential reaction?

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Bach's strategy forces him generally to distinguish between what speakers standardly/primarily mean in uttering sentences containing referring expressions and what those sentences mean. Although his pragmatic proposals attempt to make it plausible how this could be, it is noteworthy that Bach makes little or no attempt to explain why this should be. Why should there be so much indirection in our communicative practice? In the case of an isolated sentence ('I have nothing to wear'), there may be no serious difficulty in understanding how and why what speakers standardly/primarily mean in uttering it may differ from what the sentence literally means. But when we are concerned with such a vast and important class of sentences as those containing referring expressions, it is seriously questionable whether it so much as makes sense to suppose that speakers standardly/primarily mean something distinct from what they say.[6]

One way—doubtless the principal way—in which we evaluate proposed semantic theses is to see whether in fact the relevant expressions standardly get used the way the semantic proposal would lead us to expect. But what is the connection between meaning and use, between what we say and what we mean, that we take to underwrite our intuitions concerning the kind of use it is legitimate to expect, given a particular semantic proposal? This is an important question—especially now when we are threatened with a proliferation of attempts by philosophers of very different persuasions to exploit the semantic/pragmatic distinction in an effort to "accommodate" the prima facie embarrassments of their diverse semantic proposals. Perhaps, in the end, the main contribution of this provocative and tenaciously argued book will be the way it challenges its readers to think more seriously and clearly about what an adequate answer to this question would be.

WILLIAM W. TASCHEK

University of Michigan/Ann Arbor

[6] This is, of course, not to say that there are not plenty of exceptions. But they are exceptions. And our ability to make sense of them presupposes a competence with a standard of usage that it is difficult to imagine acquiring except in the context of a practice in which speakers did standardly say what they mean and mean what they say.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The editors cannot afford to send to the printer manuscripts that are difficult for the compositor to read or footnotes that need to be completed in proof. They ask that manuscripts submitted be prepared according to the following guidelines:

1. All material, including the quotations and footnotes—especially footnotes—must be typed, double-spaced, on one side of the paper, with ample margins left and right, top and bottom.
2. No bibliographies will be published; bibliographical material should be put into the footnotes.
3. Footnotes should be complete, according to *Journal* style (see any issue). Journal citations should include name of journal in full, volume number, issue number, month or season, year, and pages. Book references should include publisher, city of publication, and date. Footnotes should be gathered at the end of the article.
4. Quotations should be checked by the author; capitalization, italics, spelling, etc., should match the original. Corrections should not have to be made (expensively) in proof.
5. If there are diagrams, they should be provided camera-ready by the author. If there are special symbols, they should be widely spaced and clearly marked; if there are many, a list should be provided. Standard logic symbols are available, and letters of the greek and script alphabets; no boldface.
6. Computer copy can be extremely difficult to read; only letter-quality copy is acceptable. It should be double-spaced and not justified. The print should be dark enough and the letters widely enough spaced so that both letters and punctuation marks can be read easily.
7. Finally, photo copies on shiny paper are not acceptable. Ribbon copy is much preferred, but clear copies on bond paper are acceptable if necessary.

The editors are willing to read and evaluate manuscripts that do not meet these standards, provided they are genuinely legible and not unreasonably long. They ask for revision of excessively long manuscripts, and they require retyping of accepted manuscripts, if necessary, before scheduling for publication.

The *Journal of Philosophy* is published monthly, in issues that (ordinarily) contain only 56 pages of copy; this limits the length of papers the editors can normally consider, to no more than 7,500 words, or eighteen JOURNAL pages. Manuscripts very much in excess of these limits will be returned unread.

The *Journal of Philosophy* does not publish papers that are primarily historical or expository; the editors urge that such manuscripts be submitted elsewhere.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NOTE TO CONTRIBUTORS

As a monthly periodical, THE JOURNAL OF PHILOSOPHY encourages philosophical exchange and progress through prompt publication of articles of current interest. The editors now feel that they can ensure publication within six to twelve months of acceptance. They continue to be particularly interested in papers that explore the borderline between philosophy and the special disciplines, but are unlikely to accept articles primarily expository or historical in character. For the Comments and Criticism section they accept rejoinders but prefer those designed to advance the discussion rather than just to set the record straight. If and only if articles are held five months or more and then rejected, letters of rejection include a brief comment concerning the editors' reasons for rejection.

Attention is drawn to the memo to authors and to the statement of copyright policy, republished on page 46 and below (on this page) of this issue; copies are available from the JOURNAL office.

Contributors are supplied free of charge with ten copies of the JOURNAL issues in which their articles appear, reviewers and authors of abstracts with two copies. Offprints may be ordered from the printer on a form supplied by the printer and enclosed with galley proof; they will be sent out within six weeks of publication.

STATEMENT OF COPYRIGHT POLICY

We would like to draw your attention to the protection copyright affords against multiple photocopying by libraries, instructors, and copy centers. The 1976 copyright law provides that photocopies in excess of "fair use" copying cannot be made without authorization by the copyright owner—"fair use" meaning single copying for personal use. Accordingly, the *Journal* authorizes personal or educational multiple copying only if a request is made in writing, and the fee of $.05 per page/per copy is paid directly to us. *Each copy must include a notice of copyright.* We will turn over to the authors any substantial payments (over $20.00, an administrative fee) obtained from all photocopying. Inquiries about photocopying may be sent to the Managing Editor.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NEW BOOKS

ABEL, DONALD C.: *Freud on Instinct and Morality.* Albany: SUNY, 1989. xix, 123 p. Cloth $34.50, paper $10.95.

ALEXANDER, JEFFREY C.: *Structure and Meaning: Rethinking Classical Sociology.* New York: Columbia, 1989. viii, 258 p. Cloth $35.00.

ALICKI, ANDRZEJ: *Stanislaw Brzozowski and the Polish Beginnings of "Western Marxism."* New York: Oxford, 1989. viii, 349 p. Cloth $66.00

ALMASI, MIKLOS: *The Philosophy of Appearances.* Boston: Kluwer, 1989. xviii, 294 p. Cloth $79.00.

ALSTON, WILLIAM P.: *Divine Nature and Human Language: Essays in Philosophical Theology.* Ithaca: Cornell, 1989. xi, 279 p. Cloth $34.95, paper $12.95.

AQUILA, RICHARD E.: *Matter in Mind: A Study of Kant's Transcendental Deduction.* Bloomington: Indiana UP, 1989. xiv, 245 p. Cloth $35.00.

ARENS, KATHERINE: *Structure of Knowing: Psychologies of the Nineteenth Century.* Boston: Kluwer, 1989. x, 430 p. Cloth $84.00.

AVRAMIDES, ANITA: *Meaning and Mind: An Examination of a Gricean Account of Language.* Cambridge: MIT, 1989. xi, 202 p. Cloth $25.00.

AXINN, SIDNEY: *A Moral Military.* Philadelphia: Temple, 1989. xiv, 230 p. Cloth $29.95.

BAKER, C. EDWIN: *Human Liberty and Freedom of Speech.* New York: Oxford, 1989. vii, 385 p. Cloth $39.95.

BALLARD, EDWARD GOODWIN: *Philosophy and the Liberal Arts.* Boston: Kluwer, 1989. xvii, 342 p. Cloth $79.00.

BAUMAN, ZYGMUNT: *Modernity and the Holocaust.* Ithaca: Cornell, 1989. xiv, 224 p. Cloth $29.95.

BENARDETE, JOSE A.: *Metaphysics: The Logical Approach.* New York: Oxford, 1989. 210 p. Cloth $39.95, paper $14.95.

BENARDETE, SETH: *Socrates' Second Sailing: On Plato's* Republic. Chicago: University Press, 1989. ix, 238 p. Cloth $29.95.

BERTHOLD-BOND, DANIEL: *Hegel's Grand Synthesis: A Study of Being, Thought, and History.* Albany: SUNY, 1989. xi, 233 p. Cloth $54.50, paper $18.95.

BHASKAN, ROY: *Reclaiming Reality.* New York: Routledge, 1989. ix, 218 p. Cloth $50.00, paper $15.95.

BITTNER WISEMAN, MARY: *The Ecstasies of Roland Barthes.* New York: Routledge, 1989. xvii, 204 p. Cloth $39.95, paper $13.95.

BOUCHER, DAVID: *The Social and Political Thought of R. G. Collingwood.* New York: Cambridge, 1989. xi, 300 p. Cloth $44.50.

BRAMWELL, ANNA: *Ecology in the 20th Century: A History.* New Haven: Yale, 1989. xii, 292 p. Cloth $40.00, paper $16.95.

BRESLAUER, DANIEL S.: *Covenant and Community in Modern Judaism.* Westport: Greenwood, 1989. xi, 126 p. Cloth $35.00.

BROWN, JAMES ROBERT: *The Rational and the Social.* New York: Routledge, 1989. ix, 198. p. Cloth $45.00.

BRUNS, GERALD R.: *Heidegger's Estrangements: Language, Truth, and Po-*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

etry in the Later Writings. New Haven: Yale, 1989. xxx, 233 p. Cloth $26.50.

BURTON-ROBERTS, NOEL: *The Limits to Debate: A Revised Theory of Semantic Presupposition.* New York: Cambridge, 1989. ix, 272 p. Cloth $42.50.

CAFFENTZIS, CONSTANTINE GEORGE: *Clipped Coins, Abused Words, and Civil Government: John Locke's Philosophy of Money.* New York: Autonomedia, 1989. 246 p. Cloth $26.95, paper, $10.95.

CAMERON, SHARON: *Thinking in Henry James.* Chicago: University Press, 1989. vii, 200 p. Cloth $29.95.

CARTWRIGHT, NANCY: *Nature's Capacities and Their Measurement.* New York: Oxford, 1989. 268 p. Cloth $39.95.

CASTANEDA, HECTOR-NERI: *Thinking, Language, and Experience.* Minneapolis: Minnesota UP, 1989. ix, 302 p. Cloth $29.50.

CHISHOLM, RODERICK M.: *On Metaphysics.* Minneapolis: Minnesota UP, 1989. vii, 173 p. Cloth $29.50.

CHYTRY, JOSEF: *The Aesthetic State: A Quest in Modern German Thought.* Berkeley: California UP, 1989. xv, 513 p. Cloth $49.50.

CLARK, ANDY: *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing.* Cambridge: MIT, 1989. xiv, 226 p. Cloth $19.95.

CLAYTON, PHILIP: *Explanation from Physics to Theology: An Essay in Rationality and Religion.* New Haven: Yale, 1989, ix, 230 p. Cloth $26.50.

COCKS, JOAN: *The Oppositional Imagination: Feminism, Critique, and Political Theory.* New York: Routledge, 1989. x, 245 p. Cloth $45.00, paper $13.95.

COHEN, L. JONATHAN: *An Introduction to The Philosophy of Induction and Probability.* New York: Oxford, 1989. x, 217 p. Cloth $49.95, paper $16.95.

CONNELL, RICHARD J.: *The Empirical Intelligence—The Human Empirical Mode: Philosophy as Originating in Experience.* Lewiston, NY: Mellen, 1988. iv, 285 p. Cloth $59.95.

CREEGAN, CHARLES L.: *Wittgenstein and Kierkegaard: Religion, Individuality, and Philosophical Method.* New York: Routledge, 1989. v, 158 p. Cloth $45.00.

CROWTHER, PAUL: *The Kantian Sublime: From Morality to Art.* New York: Oxford, 1989. 178 p. Cloth $39.95.

CURRIE, GREGORY: *An Ontology of Art.* New York: St. Martin's, 1989. xii, 143 p. Cloth $39.95.

DAHL, ROBERT A.: *Democracy and Its Critics.* New Haven: Yale, 1989. viii, 397 p. Cloth $29.95.

DALFERTH, INGOLF U.: *Theology and Philosophy.* New York: Blackwell, 1988. x, 236 p. Cloth $39.95.

DANIELS, NORMAN: *Thomas Reid's 'Inquiry': The Geometry of Visibles and the Case for Realism.* Stanford: University Press, 1989. xix, 160 p. Cloth $35.00, paper $10.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

DAVIS, CAROLINE FRANKS: *The Evidential Force of Religious Experience.* New York: Oxford, 1989. ix, 268 p. Cloth $49.95.

DEL CARO, ADRIAN: *Nietzsche contra Nietzsche: Creativity and the Anti-Romantic.* Baton Rouge: Louisiana State UP, 1989. 314 p. Cloth $37.50.

DELUE, STEVEN M.: *Political Obligation in a Liberal State.* Albany: SUNY, 1989. xiv, 179 p. Cloth $39.50, paper $12.95.

DI GIOVANNA, JOSEPH J.: *Linguistic Phenomenology: Philosophical Method in J. L. Austin.* New York: Lang, 1989. xiii, 211 p. Cloth $33.70.

DILWORTH, DAVID A.: *Philosophy in World Perspective: A Comparative Hermeneutic of the Major Theories.* New Haven: Yale, 1989. viii, 232 p. Cloth $26.50.

DOBSON, ANDREW: *An Introduction to the Politics and Philosophy of Jose Ortega y Gasset.* New York: Cambridge, 1989. xiii, 186 p. Cloth $39.50.

DONALDSON, PETER S.: *Machiavelli and Mystery of State.* New York: Cambridge, 1988. xiv, 227 p. Cloth $32.50.

DUNAYEVSKAYA, RAYA: *Marxism and Freedom from 1776 until Today.* New York: Columbia, 1989. xiii, 388 p. Cloth $50.00, paper $17.50.

EIDELBERG, PAUL: *Beyond the Secular Mind: A Judaic Response to the Problems of Modernity.* New York: Greenwood, 1989. xvii, 177 p. Cloth $39.95.

EITZ, CHARLES R.: *Political Equality.* Princeton: University Press, 1989. xviii, 253 p. Cloth $25.00.

ELSTER, JON: *The Cement of Society: A Study of Social Order.* New York: Cambridge, 1989. viii, 311 p. Cloth $44.50, paper $15.95. *Nuts and Bolts for the Social Sciences.* New York: Cambridge, 1989. viii, 184 p. Cloth $34.50, paper $9.95. *Solomonic Judgements: Studies in the Limitations of Rationality.* New York: Cambridge, 1989. ix, 232 p. Cloth $39.50, paper $12.95.

FAYE, JAN: *The Reality of the Future: An Essay on Time, Causation and Backward Causation.* Odense: University Press, 1989. 321 p.

FIELD, HARTRY: *Realism, Mathematics and Modality.* New York: Blackwell, 1989. viii, 290 p. Cloth $49.95.

FISCHER, MICHAEL: *Stanley Cavell and Literary Skepticism.* Chicago: University Press, 1989. xiii, 165 p. Cloth $27.50, paper $10.95.

FLEW, ANTHONY: *Equality in Liberty and Justice.* New York: Routledge, 1989. 224 p. Cloth $49.95.

FORBES, GRAEME: *Languages of Possibility: An Essay in Philosophical Logic.* New York: Blackwell, 1989. viii, 181 p. Cloth $45.00.

FORREST, PETER: *Quantum Metaphysics.* New York: Blackwell, 1988. xvii, 238 p. Cloth $39.95.

FORRESTER, JAMES W.: *Why You Should: The Pragmatics of Deontic Speech.* Hanover, NH: New England, 1989. x, 246 p. Cloth $28.00.

FORSTER, MICHAEL N.: *Hegel and Skepticism.* Cambridge: Harvard, 1989. 256 p. Cloth $29.50.

FULLER, STEVE: *Philosophy of Science and Its Discontents.* Boulder: Westview, 1989. x, 188 p. Cloth $32.95.

GERASSI, JOHN: *Jean-Paul Sartre: Hated Conscience of His Century, vol. 1.* Chicago: University Press, 1989. ix, 213 p. Cloth $19.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

GILBERT, MARGARET: *On Social Facts*. New York: Routledge, 1989. x, 521 p. Cloth $95.00.

GIVON, T.: *Mind, Code and Context*. Hillsdale, NJ: Erlbaum, 1989. xvi, 456 p. Cloth $69.95, paper $29.95.

GOLOMB, JACOB: *Nietzsche's Enticing Psychology of Power*. Ames: Iowa State UP, 1989. 350 p. Cloth $24.95.

GORDON, HAIM: *Make Room for Dreams: Spiritual Challenges to Zionism*. New York: Greenwood, 1989. xi, 177 p. Cloth $37.95.

GRAY, ROCKWEL: *The Imperative of Modernity: An Intellectual Biography of Jose Ortega y Gasset*. Berkeley: California UP, 1989. xiii, 424 p. Cloth $29.95.

GREENWALT, KENT: *Speech, Crime, and the Uses of Language*. New York: Oxford, 1989. viii, 349 p. Cloth $45.00.

GRUMLEY, JOHN E.: *History and Totality: Radical Historicism from Hegel to Foucault*. New York: Routledge, 1989. xii, 241 p. Cloth $37.50.

GUTTING, GARY: *Michel Foucault's Archaeology of Scientific Reason*. New York: Cambridge, 1989. xiii, 306 p. Cloth $39.50, paper $12.95.

HALFON, MARK S.: *Integrity: A Philosophical Inquiry*. Philadelphia: Temple, 1989. xv, 190 p. Cloth $27.95.

HAMPSHIRE, STUART: *Innocence and Experience*. Cambridge: Harvard, 1989. 195 p. Cloth $20.00.

HANS, JAMES S.: *The Question of Value: Thinking through Nietzsche, Heidegger, and Freud*. Carbondale: Southern Illinois UP, 1989. xi, 204 p. Cloth $24.95.

HARE, R. M.: *Essays in Ethical Theory*. New York: Oxford, 1989. vii, 262 p. Cloth $39.95. *Essays on Political Morality*. New York: Oxford, 1989. vi, 264 p. Cloth $39.95.

HARRIS, ROY AND TAYLOR, TALBOT J.: *Landmarks in Linguistic Thought: The Western Tradition From Socrates to Saussure*. New York: Routledge, 1989. xviii, 199 p. Cloth $42.50, paper $14.95.

HARTLE, ANTHONY E.: *Moral Issues in Military Decision Making*. Lawrence: Kansas UP, 1989. 180 p. Cloth $25.00.

HASKER, WILLIAM: *God, Time, and Knowledge*. Ithaca: Cornell, 1989. xi, 209 p. Cloth $24.95.

HAUSMAN, CARL R.: *Metaphor and Art: Interactionism and References in the Verbal and Nonverbal Arts*. New York: Cambridge, 1989. xi, 238 p. Cloth $39.50.

HELM, PAUL: *The Eternal God: A Study of God without Time*. New York: Oxford, 1988. xv, 230 p. Cloth $49.95.

HENNINGER, MARK G.: *Relations: Medieval Theories 1250–1325*. New York: Oxford, 1989. 198 p. Cloth $39.95.

HERZOG, DON: *Happy Slaves: A Critique of Consent Theory*. Chicago: University Press, 1989. xv, 293 p. Cloth $45.00, paper $12.95.

HICK, JOHN: *An Interpretation of Religion: Human Responses to the Transcendent*. New Haven: Yale, 1989. xv, 412 p. Cloth $35.00.

HOFFMAN, *Violence in Modern Philosophy*. Chicago: University Press, 1989. ix, 163 p. Cloth $24.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

HUGHES, R. I. G.: *The Structure and Interpretation of Quantum Mechanics.* Cambridge: Harvard, 1989. xiii, 369 p. Cloth $39.50.

HUNDERT, EDWARD M.: *Philosophy, Psychiatry and Neuroscience: Three Approaches to the Mind: A Synthetic Analysis of the Varieties of Human Experience.* New York: Oxford, 1989. xiii, 346 p. Cloth $39.95.

HYMAN, JOHN: *The Imitation of Nature.* New York: Blackwell, 1989. xvii, 125 p. Cloth $34.95.

IRWIN, T. H.: *Aristotle's First Principles.* New York: Oxford, 1988. xv, 702 p. Cloth $75.00.

JANAWAY, CHRISTOPHER: *Self and World in Schopenhauer's Philosophy.* New York: Oxford, 1989. xii, 378 p. Cloth $69.00.

JOHANSSON, INGVAR: *An Inquiry into the Categories of Nature, Man and Society.* New York: Routledge, 1989. vi, 377 p. Cloth $57.50.

KAFKA, JOHN S.: *Multiple Realities in Clinical Practice.* New Haven: Yale, 1989. xi, 199 p. Cloth $22.50.

KAIN, PHILIP J.: *Marx and Ethics.* New York: Oxford, 1989. x, 222 p. Cloth $35.00.

KANERVA, PENTTI: *Sparse Distributed Memory.* Cambridge: MIT, 1988. xxxii, 155 p. Cloth $24.95.

KEEKES, JOHN: *Moral Tradition and Individuality.* Princeton: University Press, 1989. xii, 245 p. Cloth $27.50.

KEIL, FRANK C.: *Concepts, Kinds, and Cognitive Development.* Cambridge: MIT, 1989. xv, 328 p. Cloth $29.95.

KERSTEN, FRED: *Phenomenological Method: Theory and Practice.* Boston: Kluwer, 1989. x, 433 p. Cloth $98.00.

KRATOCHWIL, FRIEDRICH V.: *Rules, Norms, and Decisions: On the Conditions of Practical and Legal Reasoning in International Relations and Domestic Affairs.* New York: Cambridge, 1989. x, 317 p. Cloth $49.50.

KRIEGER, LEONARD: *Time's Reasons: Philosophies of History Old and New.* Chicago: University Press, 1989. xii, 202 p. Cloth $32.50.

KUSCH, MARTIN: *Language as Calculus vs. Language as Universal Medium: A Study in Husserl, Heidegger and Gadamer.* Boston: Kluwer, 1989. xi, 362 p. Cloth $79.00.

LACEY, A. R.: *Bergson.* New York: Routledge, 1989. x, 241 p. Cloth $39.95.

LANDESMAN, CHARLES: *Color and Consciousness: An Essay in Metaphysics.* Philadelphia: Temple, 1989. xi, 135 p. Cloth $24.95.

LANGER, MONIDA M.: *Merleau-Ponty's Phenomenology of Perception: A Guide and Commentary.* Tallahassee: Florida State UP, 1989. xix, 184 p. Cloth, $25.00, paper $16.95.

LEVIN, DAVID MICHAEL: *The Listening Self: Personal Growth, Social Change and the Closure of Metaphysics.* New York: Routledge, 1989. xxvii, 333 p. Cloth $47.50, paper $16.95.

LEVY, ZE'EV: *David Baumgardt and Ethical Hedonism.* Hoboken, NJ: Ktav, 1989. ix, 248 p. Cloth $39.50.

LINGIS, ALPHONSO: *Deathbound Subjectivity.* Bloomington: Indiana UP, 1989. 211 p. Cloth $32.50.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

LITTLE, DANIEL: *Understanding Peasant China: Case Studies in the Philosophy of Social Science.* New Haven: Yale, 1989. xi, 322 p. Cloth $30.00.

MACHAN, TIBOR: *Individuals and Their Rights.* Peru, IL: Open Court, 1989. xi, 250 p. Cloth $32.95, paper $16.95. *The Moral Case for the Free Market Economy.* Lewiston, NY: Mellen, 1988. 140 p. Cloth $39.95.

MARGOLIS, JOSEPH: *The Persistence of Reality: Texts without Referents: Reconciling Science and Narrative.* New York: Blackwell, 1989. xxiii, 386 p. Cloth $45.00.

MARRES, RENE: *In Defense of Mentalism: A Critical Review of the Philosophy of Mind.* Amsterdam: Rodopoi, 1989. 205 p.

MARTIN, RAYMOND: *The Past within Us: An Empirical Approach to Philosophy of History.* Princeton: University Press, 1989. xiii, 163 p. Cloth $25.00.

MASTERS, ROGER D.: *The Nature of Politics.* New Haven: Yale, 1989. xvii, 298 p. Cloth $25.00.

McFALL, LYNN: *Happiness.* New York: Lang, 1989. 129 p. Cloth $27.50.

McGINN, MARIE: *Sense and Certainty: A Dissolution of Scepticism.* New York: Blackwell, 1989. 166 p. Cloth $34.95.

MCNAUGHTON, DAVID: *Moral Vision: An Introduction to Ethics.* New York: Blackwell, 1988. viii, 214 p. Cloth $55.00.

MEYERING, THEO C.: *Historical Roots of Cognitive Science: The Rise of a Cognitive Theory of Perception from Antiquity to the Nineteenth Century.* Boston: Kluwer, 1989. xix, 250 p. Cloth $69.00.

MEYERS, DIANA T.: *Self, Society, and Personal Choice.* New York: Columbia, 1989. xiii, 287 p. Cloth $35.00.

MELNICK, ARTHUR: *Space, Time and Thought in Kant.* Boston: Kluwer, 1989. vii, 561 p. Cloth $99.00.

MIDGLEY, MARY: *Wisdom Information & Wonder: What is Knowledge For?* New York: Routledge, 1989. x, 275 p. Cloth $27.50.

MILLER, JEROME A.: *The Way of Suffering: A Geography of Crisis.* Washington: Georgetown, 1988. xii, 203 p. Cloth $22.95, $10.95.

MOORE, RALPH J., and BROOKE N. MOORE: *The Cosmos, God and Philosophy.* New York: Lang, 1988. 312 p. Cloth $41.50.

MOORE, WALTER: *Schrödinger: Life and Thought.* New York: Cambridge, 1989. xi, 513 p. Cloth $39.50.

MORAN, DERMOT: *The Philosophy of John Scottus Eriugena: A Study of Idealism in the Middle Ages.* New York: Cambridge, 1989. xvii, 333 p. Cloth $59.50.

MULLER, G. H.: *Sociology and Ontology: The Analytical Foundations of Sociological Theory.* Lanham: UP of America, 1989. xiv, 288 p. Cloth $28.50, paper $15.25.

NARVESON, JAN: *The Libertarian Idea.* Philadelphia: Temple, 1988. xiv, 367 p. Cloth $34.95.

NEILSEN, KAI: *God, Scepticism and Modernity.* Ottawa: University Press, 1989. 252 p.

NELSON, R. J.: *The Logic of Mind.* Boston: Kluwer, 1989. xvi, 393 p. Cloth $99.00.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NEUSNER, JACOB, et al. eds: *Religion, Science, and Magic: In Concert and In Conflict.* New York: Oxford, 1989. xii, 294 p. Cloth $29.95.

NEVILLE, ROBERT CUMMINGS: *Recovery of the Measure: Interpretation and Nature.* Albany: SUNY, 1989. xiii, 369 p. Cloth $65.50, paper $21.95.

NIELSEN, KAI: *Why Be Moral?* Buffalo: Prometheus, 1989. 300 p. Cloth $16.95.

NODDINGS, NEL: *Women and Evil.* Berkeley: California UP, 1989. ix, 284 p. Cloth $25.00.

O'HEAR, ANTHONY: *Introduction to the Philosophy of Science.* New York: Oxford, 1989. x, 239 p. Cloth $39.95, paper $10.95.

O'MEARA, DOMINIC J.: *Pythagoras Revived: Mathematics and Philosophy in Late Antiquity.* New York: Oxford, 1989. xii, 251 p. Cloth $49.95.

OPPENHEIMER, PAUL: *The Birth of the Modern Mind: Self, Consciousness, and the Invention of the Sonnet.* New York: Oxford, 1989. x, 206 p. Cloth $29.95.

OSTERBERG, DAG: *Metasociology: An Inquiry into the Origins and Validity of Social Thought.* New York: Oxford, 1988. 271 p. Cloth $39.95.

OUTRAM, DORINDA: *The Body and the French Revolution: Sex, Class and Political Culture.* New Haven: Yale, 1989. ix, 197 p. Cloth $25.00.

PAILIN, DAVID A.: *God and the Processes of Reality: Foundations of a Credible Theism.* New York: Routledge, 1989. x, 235 p. Cloth $42.50.

PALMER, MICHAEL D.: *Names, Reference and Correctness in Plato's Cratylus.* New York: Lang, 1989. xix, 207 p. Cloth $34.50.

PHILLIPSON, MICHAEL: *In Modernity's Wake: The Ameurunculus Letters.* New York: Routledge, 1989. xxiii, 174 p. Paper $9.95.

PHIPPS, WILLIAM E.: *Genesis and Gender: Biblical Myths of Sexuality and their Cultural Impact.* New York: Praeger, 1989. xvi, 125 p. Cloth $32.95.

PIPPIN, ROBERT B.: *Hegel's Idealism: The Satisfactions of Self-Consciousness.* New York: Cambridge, 1989. ix, 327 p. Cloth $44.50, paper $16.95.

PRICE, A. W.: *Love and Friendship in Plato and Aristotle.* New York: Oxford, 1989. xiv, 264 p. Cloth $45.00.

PROUST, JOËLLE: *Questions of Form: Logic and the Analytical Proposition from Kant to Carnap.* Minneapolis: Minnesota, 1989. xxii, 302 p. Cloth $35.00, paper $15.95.

RABINOW, PAUL: *French Modern: Norms and Forms of the Social Environment.* Cambridge: MIT, 1989. x, 447 p. Cloth $35.00.

RADNER, DAISIE, and MICHAEL RADNER: *Animal Consciousness.* Buffalo: Prometheus, 1989. 251 p. Cloth $39.95.

RAPOSA, MICHAEL: *Peirce's Philosophy of Religion.* Bloomington: Indiana UP, 1989. x, 180 p. Cloth $25.00.

RESCHER, NICHOLAS: *Moral Absolutes: An Essay on the Nature and Rationale of Morality.* New York: Lang, 1989. x, 115 p. Cloth $29.00.

RICKHOUSE, THOMAS C., and NICHOLAS D. SMITH: *Socrates on Trial.* Princeton: University Press, 1989. xi, 337 p. Cloth $35.00.

ROCKMORE, TOM.: *Habermas on Historical Materialism.* Bloomington: Indiana UP, 1989. xii, 202 p. Cloth $35.00, paper $14.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ROLLIN, BERNARD E.: *The Unheeded Cry: Animal Consciousness, Animal Pain and Science.* New York: Oxford, 1989. xviii, 308 p. Cloth $29.95.

ROLLINS, MARK: *Mental Imagery: On the Limits of Cognitive Science.* New Haven: Yale, 1989. xvii, 170 p. Cloth $21.50.

ROLSTON, HOLMES, III: *Philosophy Gone Wild: Environmental Ethics.* Buffalo: Prometheus, 1989. 269 p. Paper $14.95.

ROSS, STEPHEN DAVID: *Metaphysical Aporia and Philosophical Heresy.* Albany: SUNY, 1989. viii, 411 p. Cloth $73.50, paper $24.50.

RUSE, MICHAEL: *The Darwinian Paradigm: Essays on its History, Philosophy and Religious Implications.* New York: Routledge, 1989. ix, 299 p. Cloth $25.00.

RUSSOW, LILLY-MARLENE, and MARTIN CURD: *Principles of Reasoning.* New York: St. Martin's, 1988. xi, 372 p. Paper $22.00.

SAMUELS, ANDREW: *The Plural Psyche: Personality, Morality, and the Father.* New York: Routledge, 1989. xiv, 253 p. Cloth $25.00.

SCHARFSTEIN, BEN-AMI: *Of Birds, Beasts, and Other Artists: An Essay on the Universality of Art.* New York: University Press, 1988. xi, 313 p. Cloth $19.95. *The Dilemma of Context.* New York: University Press, 1989. xv, 218 p. Cloth $35.00.

SCHRAM, STUART: *The Thought of Mao Tse-Tung.* New York: Cambridge, 1989. x, 242 p. Cloth $44.50, paper $14.95.

SHAPIRO, GARY: *Nietzschean Narratives.* Bloomington: Indiana University, 1989. 179 p. Cloth $37.50, paper $14.95.

SIDELLE, ALAN: *Necessity, Essence, and Individuation: A Defense of Conventionalism.* Ithaca: Cornell, 1989. xv, 216 p. Cloth $26.50.

SILVERMAN, IRVING: *The Supply of Concepts.* New York: Praeger, 1989. xv, 182 p. Cloth $39.95.

SLOTE, MICHAEL: *Beyond Optimizing: A Study of Rational Choice.* Cambridge: Harvard, 1989. viii, 192 p. Cloth $22.50.

SMYTHIES, JOHN R., and JOHN BELOFF: *The Case for Dualism.* Charlottesville: Virginia UP, 1989. xii, 268 p. Cloth $35.00.

SOBER, ELLIOT: *Reconstructing the Past: Parsimony, Evolution and Inference.* Cambrige: MIT, 1989. xii, 265 p. Cloth $27.50.

SONTAG, FREDERICK: *Emotion: Its Role in Understanding and Decision.* New York: Lang, 1989. 227 p. Cloth $35.95.

STEINER, GEORGE: *Real Presences.* Chicago: University Press, 1989. 236 p. Cloth $19.95.

STORR, ANTHONY: *Freud.* New York: Oxford, 1989. 135 p. Cloth $27.95, paper $5.95.

SULLIVAN, ROGER J.: *Immanuel Kant's Moral Theory.* New York: Cambridge, 1989. xvii, 413 p. Cloth $49.50, paper $16.95.

SUPPE, FREDERICK: *The Semantic Conception of Theories and Scientific Realism.* Chicago: Illinois UP, 1989. xvi, 475 p. Cloth $34.95.

SWANSON, JENNY: *John of Wales: A Study of the Works and Ideas of a Thirteenth Century Friar.* New York: Cambridge, 1989. xi, 307 p. Cloth $49.50.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

SWINBURNE, RICHARD: *Miracles*. New York: Macmillan, 1988. 210 p. Paper $9.00.

TAYLOR, CHARLES: *Sources of the Self: The Making of the Modern Identity*. Cambridge: Harvard, 1989. xii, 601 p. Cloth $37.50.

THOMPSON, PAUL: *The Structure of Biological Theories*. Albany: SUNY, 1989. x, 148 p. Cloth $39.50, paper $12.95.

TIESZEN, TICHARD L.: *Mathematical Intuition: Phenomenology and Mathematical Knowledge*. Boston: Kluwer, 1989. xiv, 209 p. Cloth $69.00.

TRAVIS, CHARLES: *The Uses of Sense: Wittgenstein's Philosophy of Language*. New York: Oxford, 1989. xiii 400 p. Cloth $55.00.

TU, WEI-MING: *Centrality and Commonality: An Essay on Confucian Religiousness*. Albany: SUNY, 1989, x, 165 p. Cloth $39.50, paper $12.95.

VICKERS, JOHN M.: *Chance and Structure: An Essay on the Logical Foundations of Probability*. New York: Oxford, 1988. viii, 244 p. Cloth $52.00.

VISION, GERALD: *Modern Anti-Realism and Manufactured Truth*. New York: Routledge, 1989. xviii, 308 p. Cloth $65.00.

VOEGELIN, ERIC: *Autobiographical Reflections*. Baton Rouge: Louisiana State UP, 1989. xiii, 123 p. Cloth $16.95.

WALDRON, JEREMY: *The Right to Private Property*. New York: Oxford, 1988. viii, 470 p. Cloth $59.00.

WALTON, DOUGLAS N.: *Informal Logic: A Handbook for Critical Argumentation*. New York: Cambridge, 1989. xiii, 292 p. Cloth $44.50, paper $12.95.

WARNOCK, G. J.: *J. L. Austin*. New York: Routledge, 1989. 164 p. Cloth $33.00.

WEIGEL, VAN B.: *A Unified Theory of Global Development*. New York: Praeger, 1989. xix, 263 p. Cloth $47.95.

WESSON, ROBERT: *Cosmos and Metacosmos*. La Salle, IL: Open Court, 1989. xii, 157 p. Cloth $26.95, paper $11.95.

WESTPHAL, KENNETH R.: *Hegel's Epistemological Realism: A Study of Aim and Method of Hegel's Phenomenology of Spirit*. Boston: Kluwer, 1989. xiv, 309 p. Cloth $69.00.

WIERENGA, EDWARD R.: *The Nature of God: An Inquiry into Divine Attributes*. Ithaca: Cornell, 1989. ix, 238 p. Cloth $28.50.

WILLIAMS, PAUL: *Mahayana Buddhism: The Doctrinal Foundations*. New York: Routledge, 1989. xii, 317 p. Cloth $55.00, paper $17.95.

WILSON, GEORGE M.: *The Intentionality of Human Action*. Stanford: University Press, 1989. xi, 299 p. Cloth $34.50.

WITT, CHARLOTTE: *Substance and Essence in Aristotle: An Interpretation of* Metaphysics VII–IX. Ithaca: Cornell, 1989. viii, 201 p. Cloth $24.50.

WOLFE, ALAN: *Whose Keeper? Social Science and Moral Obligation*. Berkeley: California UP, 1989. xvii, 371 p. Cloth $25.00.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVII, NO. 2, FEBRUARY 1990

IN DEFENSE OF '⊃'*

What, after all, *is* the relationship between the English word 'if' and the horseshoe operator of elementary logic? If we regard such conditionals as 'if Tom marries Joan, Susan will be unhappy', or 'if you drop that, it will break', as truth functional, is this justified, or is it, as some philosophers have said, definitely wrong?

In elementary logic, for the purpose of analyzing arguments, etc., we do regard such statements as these as truth functions of their components; we say that their truth values depend on the truth values of the components and on nothing else. So, for example, the first statement I mentioned is counted as true if Tom will not marry Joan, whether or not Susan will be unhappy, as true if Susan will be unhappy, whether or not Tom will have married Joan, and as false in the one case in which Tom will marry Joan and Susan will not be unhappy. Now, it is possible to hold that in logic we adopt this interpretation of conditionals merely for convenience, because of certain given purposes, so that the view is a kind of crude approximation to the facts. Then someone would be entitled to use truth-functional logic and make certain claims on its behalf without being committed to the view of conditionals in question being anything more than crude approximation. But one can still ask: If it is nothing more than an approximation, how good a one is it? And one can raise the same question in a more dramatic way by asking: If someone does hold that the view is perfectly correct, is he right or wrong? I think it is no exaggeration to say that practically everyone who has discussed

* This essay was one of the late James F. Thomson's unpublished papers. We believe that he wrote it in 1963 or 1964. We are grateful to Carolyn Farrow for transcribing the manuscript, and to Robert Stalnaker and James Higginbotham for helpful comments; we have made only minor editorial changes in preparing the essay for publication. [Note by George Boolos and Judith J. Thomson.]

0022-362X/90/8702/57–70

© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the question at any length has said that this would be definitely and flatly wrong, a quite definite mistake.

So many writers have held this view—from the late G. E. Moore and P. F. Strawson, down to my very youngest student—that one cannot help feeling that there must be something in it, or at least something behind it. But I am afraid that it is not clear to me just what there is in it, or behind it. So, without pretending to have a clear view of the matter, I would like to ask: Have any really conclusive reasons been advanced for thinking that conditionals of our type are not truth functional?

I say 'of our type', because I am anxious to restrict the discussion to some conditionals, namely, those of which it is at least plausible to hold that they are truth-functional. I am sure that many conditionals are not truth-functional. I shall not try to offer an exact delimitation of the ones to which I want my discussion to be relevant. But I explicitly exclude contrary-to-fact conditionals; also such conditionals as if true are logically or analytically true; and also those which J. L. Austin recently noticed, such as 'If you're hungry, there are some biscuits on the sideboard'. These latter have the peculiarity of not implying their contrapositives. Someone who says, 'If you're hungry, there are some biscuits on the sideboard', is not committed to saying that, if there are not any biscuits there, then his hearer is not hungry. This will have to do as a rough indication of the conditionals about which I do want to ask. They, I think, have the best claim of any to be truth functions if any have; that they have any claim at all has certainly been denied; and I want to ask whether any conclusive reasons for that denial have been put forward.

Let me start from this point: on the truth-functional account of a conditional, each of the following statements is true:

If Napoleon is alive, Oxford is in France.
If Napoleon is dead, Oxford is in England.
If Napoleon is alive, Oxford is in England.

For each has either a false antecedent or a true consequent; the third of course has both. Now, these statements are admittedly rather odd. This has to be admitted by anyone. But what makes them odd? It is not of course enough to exclaim, 'But we just wouldn't say anything like that'—perhaps we wouldn't—but why?

One answer to our question is offered by what I should like to call the *received opinion* (RO) about conditionals. This is, roughly, to the effect that a statement 'if p then q' is stronger than the corresponding material conditional '$p \supset q$'. 'If p then q' implies '$p \supset q$', but is not implied by it. In other words, for it to be true that if p then q, it has to

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

be true that p materially imply q, but this, though necessary, is not sufficient. What else is then required? Well, this extra necessary condition is variously described by those who hold what I am calling the RO. Sometimes they say that there must be some or other kind of connection between the antecedent and the consequent, or that the antecedent must somehow be relevant to the consequent. A more exact statement of the requirement is sometimes given in terms of the idea of a ground or reason. According to this, the antecedent must be so related to the consequent that the truth of the antecedent is or would be a reason for accepting the consequent. This then imposes a test on conditionals. And it is clear, or at least seems clear, that none of our three deviant conditionals passes the test. Someone who thought that Napoleon was still alive would not thereby have a reason for thinking that Oxford was in France; nor, for that matter, that it was in England. And those of us who suppose Napoleon long dead do not count his being so as among our reasons for thinking that Oxford is in fact in England.

So the received opinion does then supply us with an answer to our question. But let us notice carefully what that answer comes to. This RO says that such-and-such is a necessary condition for the truth of a conditional statement. Our 'odd' conditionals do not satisfy that condition. So, on any ordinary view of statements, it seems that someone who adopts that view is committed to saying that our odd conditionals are false. And this could be one explanation of why they do seem odd to us. But now is that right? Will it do?

To me at least it seems definitely wrong. The RO takes its stand on an appeal to ordinary language, to our ordinarily unformulated intuitions as to what we would and would not ordinarily say, being in possession of the relevant facts, and to what, being in possession of them, we would or would not call true. Now, it may well be that we would not ordinarily want to call the three odd conditionals 'true'. About this for the moment I express no opinion. But would we want either to call them false? It seems to me that anyone who hesitated at calling them true would hesitate equally to call them false. If you gave a list of conditionals to someone uninterested in and ignorant of logic and asked him to mark them T, F, or ?, wouldn't he mark our three examples as '?'? I think he would.

And if this is correct, then the answer given by the RO is just mistaken. It follows too, of course, that the RO is itself mistaken. It claims to offer an account of the relationship between 'if p then q' and '$p \supset q$'. But it will now seem that the account it gives is mistaken.

In case this victory seemed too easy, let me interpolate two remarks. An example of a "mode of statement composition" that ev-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

eryone would agree to be non-truth-functional is that afforded by the schema 'p, and that because q'. For it to be true that Susan is unhappy and is unhappy *because* Tom has married Joan, it must be true that Susan is unhappy, and it must be true too that Tom *has* married Joan. But these necessary conditions are plainly not sufficient, since it might be true that Susan is unhappy not because of Tom—even in fact *despite* Tom having married Joan. So here we can say that 'p and that because q' implies but is not implied by 'p and q'. Furthermore, since the extra condition required by the former type of statement is not anything to do with the truth values of p and q (it may of course have something to do with the truth values of some other statement or statements), this really does make it clear that 'p and that because q' is not a truth function of p and q. And so it is at least clear that, if anyone holds that 'if p then q' is non-truth-functional in much the same way as 'p and that because q', he is just wrong about the linguistic facts. For if Susan is unhappy despite Tom's having married Joan, the statement that she is unhappy *because* Tom married Joan is flatly and definitely false. But if Tom does not marry Joan, it is not in the same way or to the same extent flatly false to have said, 'If Tom marries Joan, Susan will be unhappy'. So I can at least claim that the RO owes us an explanation of the kind of non-truth-functionality that is, according to it, exemplified by 'if p then q'.

The second remark is this. You may feel that I am taking the RO in too down-to-earth a way. Perhaps I am—I admit it. I am supposing that, if someone says that 'if p then q' implies, but is not implied by, '$p \supset q$', he really does mean that the inference from '$p \supset q$' to 'if p then q' is in general fallacious; and I suppose that, if this is so, there have got to be values of 'p' and 'q' for which the inference fails, i.e., values which make '$p \supset q$' true and 'if p then q' false (at least, on any conventional view of statements, validity and nonvalidity). And of course I am supposing, too, that our examples supply such values—but if these do not I am not sure what examples would. And if to suppose this is to be unsympathetic, then I must plead guilty of the crime.

But anyway, if this is not what the RO wants to say, what does it want to say? There is a temptation here to fall back on saying: despite what you have said, '$p \supset q$' does fail to imply 'if p then q', but in the following sense, namely, that the former might be true, whereas the latter is, not quite false, but all the same something we should not be willing to assert. But, in the first place, there is surely no such sense of 'implies' as that. To say that one statement does not imply another is surely to say that anyone who claimed to be able to deduce the other from the one would be arguing fallaciously. Consider: if you

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

say that I owe you five pounds, and I say, 'so it follows then that I owe you 2400 halfpennies', then my saying that may well be out of place, pointless, anything you like. But it is one thing to say that we would not "ordinarily say that," and quite another thing to object that it does not follow, i.e., that I have *made a mistake in my arithmetic.* And, in the second place, to fall back on inventing a sense of 'entails' or 'implies' is only to leave the original problem where it was. For it has already been conceded that there is something queer about our three examples of conditionals. We are trying to find out just *what.* And so it should still seem that the RO either gives the wrong answer to our question or gives none at all, but simply restates the problem.

But at this point a way of reinterpreting the RO will I think occur to us. Let us agree that our examples are not naturally called 't' or 'f'. Then they are so far indeterminate. And let us take the RO as laying down, not a necessary condition for the truth of a conditional, but a necessary condition for its being determinate. Call this the RO, second version.

Strawson,[1] who quite explicitly says that 'if p then q' implies but is not implied by '$p \supset q$', and so holds what I called RO, actually seems to me to vacillate between RO and something like this second version. He says:

> The standard or primary use of an 'if . . . then . . .' sentence, on the other hand, we saw to be in circumstances where, not knowing whether some statement which could be made by the use of a sentence corresponding in a certain way to the first clause of the hypothetical is true or not, or believing it to be false, we nevertheless consider that a step in reasoning from that statement to a statement related in a similar way to the second clause would be a sound or reasonable step; the second statement also being one of whose truth we are in doubt, or which we believe to be false. Even in such circumstances as these we may sometimes hesitate to apply the word 'true' to hypothetical statements (i.e. statements which could be made by the use of 'if . . . then . . .' in its standard significance), preferring to call them reasonable or well-founded; but if we apply the word 'true' to them at all, it will be in such circumstances as these. Now one of the sufficient conditions of the truth of a statement of material implication may very well be fulfilled without the conditions for the truth (or reasonableness) of the corresponding hypothetical being fulfilled; i.e., a statement of the form 'p ⊃ q' does not entail the corresponding statement of the form 'if p, then q'. But if we are prepared to accept the hypothetical statement, we must in consistency be prepared to deny the conjunction of the statement corresponding to the first clause of the sentence used to make the hypotheti-

[1] *Introduction to Logical Theory* (London: Methuen, 1952).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

cal statement with the negation of the statement corresponding to its second clause; i.e., a statement of the form 'if p, then q' does entail the corresponding statement of the form 'p ⊃ q' (*ibid.*, p. 83).

Now, I think it has to be agreed that it is not really clear what this comes to. Implicit in it are the following suggestions:

(1) We often hesitate to apply 'T' or 'F' to a conditional, and prefer to call it 'well-founded' or 'reasonable'.
(2) We shall be most ready to call a conditional "reasonable" if its antecedent does or would provide us with a reason for accepting the consequent.
(3) But, if we are going to call a conditional "true," it will be when it is reasonable in that sense.
(4) But, anyway, it is not just *any* reasonable conditional that we should even then be willing to call "true"; we should be willing to call it true only when it is made in what we might call "standard circumstances," i.e., when the speaker either does not know whether the antecedent is true or believes it false, and does not know whether the consequent is true or believes it false, and (thirdly) thinks that the truth of the antecedent would be a ground or reason for accepting the consequent.

And here several questions seemed to be mixed up. Let me try to untangle them. First, suppose someone to say, 'If Tom marries Joan, Susan will be unhappy', and suppose him to say this simply for effect, irresponsibly, and having no reason to think that Tom's marrying Joan would be likely to make Susan unhappy. Then he is not making the statement in what I called standard circumstances, i.e., he is not using his sentence in what Strawson called the primary or standard way. But suppose further that we, his hearers, happen to know that Susan is in love with Tom and will be unhappy if he marries Joan and not her. Now, does Strawson wish to hold that it would be wrong for us to say, "That's true, she certainly will," or to say, "That's a reasonable thing to suppose"? Taken in one way, what he says does suggest that he would deny this, just because the utterance of the conditional does not exemplify his conditions for the "standard use." But, of course, taken in another way, he might deny it. He might say that, if we, his hearers, repeat his statement, we shall be making the statement in standard circumstances, and will then be entitled to call our statement true, and then, if we are making the same statement, we are entitled to call the other man's statement true as well.

I submit that we ought anyway to call this other man's statement true. What he says is true, even though he has no reason (in the appropriate sense of 'reason') for saying it, i.e., he has no reason to

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

think it. Also, his statement is a reasonable one. It is not of course reasonable for him to make it. But it is a statement for which good reasons can be given. (About whether it can be called well-founded I admit doubts, but I shall not bother about this.)

Let us suppose then that Strawson would be willing to admit that the irresponsible man's statement can be called true (if of course we are prepared to call any conditional true).

But there is now another difficulty. Strawson says that it is in such "circumstances as these" that we shall be willing to apply 'true', if we are willing to apply it at all, namely, when its antecedent does or would provide us with a reason for accepting the consequent. But all that is guaranteed by our making a conditional statement in these circumstances is that we think that the step in reasoning from antecedent to consequent would be sound or reasonable. And, of course, from our thinking that it would be it hardly follows that it is. And so, when he speaks of our thinking that the step from p to q would be sound or reasonable, he must mean that we shall not be making the conditional in standard circumstances unless we think that the step would be sound or reasonable *and* think this rightly. So the conditions for a step to be "sound or reasonable" must exclude this much, at least that the step is from a true statement to a false one. For otherwise, if this is not excluded, 'if p then q' might be reasonable enough, in the sense that it would be reasonable to think it yet '$p \supset q$' be false; but then it would be quite wrong on any view to say that 'if p then q' implies '$p \supset q$'. Perhaps this is intended by Strawson; perhaps that is the point of the word 'sound' in the phrase 'sound or reasonable'—a sound inference being one which is at least not from a truth to a falsehood. But an obvious question now arises. The idea was, we thought, that, 'if p then q' had to satisfy conditions of some sort which '$p \supset q$' did not need to. But those conditions cannot merely be that the step from p to q should be sound in that very weak sense, since '$p \supset q$' has to satisfy that condition, too. So, what does 'if p then q' have to be to be called reasonable over and above the step from p to q being sound in that weakest possible sense?

Now, if Strawson was trying to characterize the conditions for a conditional to "sound all right," in the way that our three odd ones do not, then obviously he has failed, since it is not a necessary condition for a conditional to sound all right that it should not be materially false. 'If Tom marries Joan, Susan will be unhappy' sounds all right and may still be materially false. (I just think that Susan is in love with Tom; she is not really.) But there is another and more serious trouble. Satisfaction of Strawson's other requirements is not sufficient for a conditional to sound all right—and, worse still, a

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

conditional may sound all wrong and (not only satisfy all Strawson's requirements but also) be all right.

Suppose, for example, that you and I are arguing about conditionals, and you say that some connection is required between antecedent and consequent, or something like that, and I deny it. I ask you to consider '*If* there is a copy of *Moby Dick* on that table, then there was at least one very large Great Dane with a solid gold collar in Paddington Station yesterday'. I ask you whether this is not true. You look, and see that there is no copy of *Moby Dick* on the table, and think (I hope) that the conditional is true, but only as a material conditional, and that, since there is no connection, it does not count as true in any other better way. I hope you say that, because I now tell you that Mary said that she might be coming to Oxford, and that if she came she would return my copy of *Moby Dick* and (if I was not in) leave it on the table; so that if there is a copy of *Moby Dick* there, Mary came from London to Oxford yesterday, so that, since she certainly came by train, she must have been in Paddington; and finally I add that Mary never travels without her dog, which is a large Great Dane, etc.

The point of this example is: it may look as though someone is asserting a conditional for no better reason than that it has a false antecedent, and yet this may not be the case at all. If we see no way in which the antecedent is relevant to the consequent, but notice that the antecedent is false, we may guess that it is being asserted just because the antecedent is false. But, my only point is, we may be wrong in our guess.

It seems, or I anyway now want to suggest, that we must distinguish between two things that Strawson confuses. There is the question when, if at all, we say that a conditional is true (perhaps in the sense of being somehow reasonable), and the question when a particular act of assertion of a conditional is both reasonable as an act and is in point. I hope to make it plausible that our odd conditionals are not odd as statements, that what is or would be odd is the act of making them, or rather, that to make them would, in a sense, be absurd. The way will then be left open for our calling them true if we so wish; I shall make a rather half-hearted recommendation in that direction.

It is sometimes said that the material truth of such a conditional as 'if Napoleon is alive then Oxford is in France' follows from the truth table for '⊃'. This is in one way false, and on any view careless. For the only statements whose truth is guaranteed by the truth table are the tautologies, and that statement is not a tautology. If we are willing to call it true (or materially true) then its truth is contingent,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

because contingent on the fact that Napoleon is dead. What we can assert on the basis of the truth table alone is:

If Napoleon is dead, then, if Napoleon is alive, then Oxford is in France.

And if we are willing to suppose it true that Napoleon is in fact dead, we can apply *modus ponens,* and so arrive at the consequent of the tautology. That truth-functional logic allows us to do that is certainly a fact. It does permit the inference from 'not-p' to 'if p then q' and from 'q' to 'if p then q'. And a supporter of the logic must then be willing to say that these inferences are valid. But he need not say that it would be reasonable actually to draw them, or to acquiesce in someone else's drawing them. Consider a parallel case (perhaps too nearly parallel to suit my purpose, but never mind): *A* says to *B*, "I heard from John; he's in Rapallo," and *B* replies, "Oh. It follows then that he's either in Rapallo or somewhere in the Shetlands." *B*'s remark will now be found puzzling and pointless. Why does he bother to say that? But to criticize his remark on such grounds as these is one thing; to say he has committed a logical mistake, argued fallaciously, is quite something else.

Let me try to bring out the oddness of this kind of argument. There is an oracle (*O*) that makes statements, and an acolyte, a recorder (*R*), who records them. *R* also makes deductions from the statements so recorded. So if *O* says on one day 'p', and on another 'if p then q', *R* writes these and also may write down 'q'. *R* allows himself just such deductions as are codified in elementary logic. I shall suppose that, if *O* seems to contradict itself, *R* notices this, blames himself, and makes such minimal alterations in his book of records as will preserve consistency. If now *O* says 'p', *R* will certainly write down 'p' and he *may* also, for all we have said, write down 'if not-p then q'. But he *will* not. For, if he writes down 'if not-p then q', he will not be able to use this to any purpose. For if *O* were to say 'not-p', *R* could not deduce 'q', but must notice the inconsistency and square his book; if he erases 'not-p', he cannot now appeal to *modus ponens,* since he lacks a necessary premise, and if he erases 'p', he loses his ground for saying 'if not-p then q', and so again lacks a necessary premise. Nor, if *O* says 'q', will he write down 'if p then q'. For if later *O* were to say 'p', there would be no point in appealing to *modus ponens* to get 'q', since *ex hypothesi* 'q' is already in *R*'s book. And if *O* were later to say 'not-q', *R* could not appeal to *modus tollens* to get 'not-p', but must as before square his book.

The point here is one made quite some time ago by W. E. Johnson, and it can be generalized. If we allow that 'if p then q' follows from

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

'not-p', we must allow, too, that whatever follows from 'if p then q' and 'not-p' together follows from 'not-p' alone. Again, if we allow that 'if p then q' follows from q, we must allow, too, that whatever follows from 'if p then q' and q together follows from q alone. This means that in general there will be nothing R might wish to deduce for the deduction of which he would need as extra premises what he here and now deduces from what O says. O, as it were, gives R all he needs; initiative is not called for in the recorder's job. Now, one purpose we have in deducing consequences from a set of premises is to see what simpler statements, statements whose content is more easily grasped, are entailed by those premises. (For of course infinitely many statements are entailed by any set; but it is not likely to be interesting that from, e.g., p we can get 'if q then p and q'. Or rather, it will be interesting if we are interested to know, in respect of some particular statement q, what we should be able to assert and are not now able to assert, if it were discovered that q.) And so, it is going to be a waste of time to infer 'if p then q' from 'not-p' or from 'q'. The logic allows it, yes. But that does not mean that we have got to do everything that the logic allows. As Ludwig Wittgenstein remarked, in a rather different connection, just because there is no fence along the top of a cliff does not mean that you have got to fall over. Not everything not excluded by the rules is thereby encouraged or required.

But this will not be thought a complete defense. It deals with the case of a man who says 'not-p', so 'if p then q'. But it does not deal with the case of a man who says 'if p then q' on the sole ground that not-p.

To try and deal with this case, let me first go back to the oracle. Let us now suppose that customers come from time to time and ask: Does O think that so-and-so? R looks through his book to see what O has said about the matter, and also what O has committed itself to. If, for example, the question is whether is it so that p and if O has not said p or said 'not-p', but has said (e.g.) r and 'if r then p', R will say, O thinks that p. Suppose now that a customer asks whether it is the case that if p then q; and R, looking through his book, finds that O has in fact said quite explicitly that q. Now what shall he say? I want to say that this is not a question of logic at all but a question of commercial policy and perhaps of commercial ethics. A generous and helpful recorder will say: "Well, as a matter of fact, he's said q outright. So, if p then q, and also if *not-p* then q, if you see what I mean." But then why should he? If the customer is paying for information, why give him more information than he is paying for? And so he may say: O thinks that if p then q, i.e., the answer to your question is "Yes."

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The point here is that q contains more information, at least in general, than 'if p then q'. And 'not-p' contains more information than 'if p then q'. (An aside: the notion of information needed here can be made quite exact if we (a) make an agreement about what is to count as an atomic statement and (b) assign a probability measure over our atomic statements. But any set of statements will do as a set of atomic ones provided they satisfy certain quite formal requirements, i.e., we are not going to be committed to any metaphysics.)

In the same way p will in general give more information than 'either p or q'. Now, in ordinary life we expect each other to be at least moderately generous with our information. So if you ask me where John is, and I say 'either in Oxford or in London', you will tend to take for granted that I am giving you as much information as I relevantly can. If you later discover that I knew at the time that John was in Oxford, and either did not want you to know that, or did not want you to know that I knew, you will think that I was mildly dishonest. For I allowed you to believe that I did not know whether he was in Oxford or not; although I was perhaps, in a casuistical kind of way, careful not to say, in so many words, 'I don't know which'. Whether this should be called lying I am not clear. But this much is obvious: if I say that John is either in Oxford or in London, and say nothing more, then, if John is in fact in Oxford, I have not said anything false, since what I do say is compatible with all the relevant facts.

Compare that case with a slightly different one. A child asks to be taken to the circus. His mother already intends to take him to the circus, but says, 'If you're good I'll take you'. This is a deception (perhaps justifiable) of much the same kind. For the child will tend to suppose that he will be taken to the circus only if he is good, whereas the fact is, we are supposing, that he will be taken whether he is good or not. Here we see clearly that the mother did not say, though she did allow and perhaps even encourage the child to believe, that if he was not good he would not be taken. So, in giving the child less information than she might have done, she may bring it to behave in a different way than it otherwise would.

I think we are now in sight of a clearer view of what is a moderately complicated situation. In saying 'if p then q' a speaker will say something which is in general anyway true or false. But by the act of making the statement he will do other things, too. He will encourage us to think that he has some or other reason for thinking that if p then q and that his reasons are not such as to allow him to assert not-p nor such as to allow him to assert q. If we take these conditions as defining a sensible assertion act, we can offer a suggestion about the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

three odd conditionals with which we started off. What is odd about them will now seem this: if someone were to assert any of them, he would not, we feel, be making a sensible assertion act. For either his reason lies in the falsity of the antecedent or in the truth of the consequent; or it does not and then we just do not see what his reason could be.

And to give this account leaves the way open to say that these conditionals are true, after all. Should we take it? We do not need to. But I think we can if we like. And we may be emboldened to take it when we notice that language already has a device that will, so to speak, ease the shock: the phrase 'trivially true'. A certain logician once made a conjecture to the effect that every sequence of rationals that satisfied a certain condition (was definable in a certain system) had a certain property. Another logician showed, quite some time later, that no sequences of that kind were definable in that system anyway. Both logicians put it that conjecture had been shown to be trivially true.

One more point. When someone makes a conditional statement, we not only tend to suppose him to have reasons of a nontrivial sort, but tend actually to impute particular reasons to him. This is, I believe, at least part of the explanation for the idea that a conditional at least typically exhibits a stronger connection between antecedent and consequent than is suggested by the truth table for $\supset$. Thus, if someone says, 'If you drop that glass it will break', we tend to suppose not just that he has a reason for thinking so but that he has this reason in particular, namely, that the glass is brittle and the floor beneath it hard. And because of that, we read what we take to be his reason into the statement itself, and behave as if he had said, 'If you drop it, it will break and will break because it is brittle'. Now this, regarded as a function of three statements—'You will drop it'. 'It will break'. 'It is brittle'.—is certainly not a truth function of them. Indeed, if we take his statement to be (roughly) 'If you drop it, it will break and will break because you dropped it', then that, though a function of 'you will drop it' and 'it will break', is not a truth function of them anymore than 'If Tom marries Joan, Susan will be both unhappy and unhappy because of Tom's marrying Joan' is a truth function of its components. No wonder then that philosophers have denied that 'if you drop it, it will break' is a truth-functional conditional.

Perhaps the easiest way of coming to see that we should not import the imputed reasons into the information conveyed by the statement is by paying more attention to cases where it is easy to go wrong over

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

one's "imputation." So, to end, let me mention a rather farfetched case of that kind.

An astronomer, living among a barbarous and savage tribe, employs the following device to maintain his prestige and keep himself out of the cooking pot. From time to time he says, 'If I clap my hands there will be an eclipse'; and when challenged, claps them. And of course there is. He is careful to say this only when he knows there will be an eclipse. Perhaps we, who also live among these people, will remonstrate with him: 'But (we say), there is going to be an eclipse whether you clap your hands or not'. He has a reply to this: 'Your point is', he says, 'that there will be an eclipse if I clap, and will be one if I don't. Quite so. So you are saying that if p then q and that if not-p then q, and I was just saying if p then q, so what I said follows from what you said, and how can it be an objection to me to say something from which what I said follows'. So we shall have to fall back on saying that he is deceiving the natives, since they come to believe that there is a causal or a magical connection between his hand-clapping and eclipses. And then of course he will have to agree that this is his intention.

Now, so far we have one and the same statement made for two different reasons. The astronomer says that if he claps there will be an eclipse, for no better reason than there will be an eclipse. The savages say that if the astronomer claps there will be an eclipse, and say this for a more interesting reason, but a superstitious one. But notice now that we, who do not know much about eclipses, but do know a lot about the astronomer, may have a third reason for saying that if he claps there will be an eclipse, and this not a superstitious one; the astronomer will not clap unless he thinks there will be an eclipse because he will not want to lose face, and he is not likely to think there will be an eclipse unless there is going to be one, because he is a good astronomer, and because he has every reason (the cooking pot) for not making mistakes. So if one of us says 'Will there be an eclipse? Yes, I think so, because I saw the astronomer clapping in the market place', he is making a sound or reasonable step, and is then entitled to say that if the astronomer claps his hands there will be an eclipse.

So we have now the phenomenon of one and the same statement made for three quite different kinds of reason: the astronomer's, cogent but dull; the savages', not dull but silly; ours, cogent, and far from silly. Plainly then, it will not do to import the reasons for which the conditional is asserted into the meaning of the conditional.

So what hope is there of being able to pick out conditionals as

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

 Digitized by Arya Samaj Foundation Chennai and eGangotri

(somehow) *specially* true, not just *materially* true, in virtue of a connection between antecedent and consequent, either of the "step in reasoning" kind or of some other kind? I hope I have made it plausible that there is not much. That is why I said at the beginning that I was not happy about the current rejections of the material conditional. For they all seem to promise an account of nonmaterial truth of a conditional which will accord better with our actual speech habits than the account provided by the truth table. I have tried to make it look as if those promises were just promissory. And then the way is left open, as I said, for calling 'true' any conditional that is just "materially true." I did not say that we had to take that way that was left open. Nor did I say that there might not be an account of 'real' conditionals which fared better than those I have discussed. But certainly I do not know of any.

JAMES F. THOMSON

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

PERSONHOOD AND PERSONAL IDENTITY*

The question 'Who am I?' might be asked either by an amnesia victim or by a confused adolescent, and requires a different answer in each of these contexts. In the former case, the questioner is asking which history her life is a continuation of, and, in the latter, the questioner presumably knows her history but is asking which of the beliefs, values, and desires that she seems to have are truly her own, expressive of who she is. These can be called, respectively, the *question of reidentification* and the *question of self-knowledge*. Contemporary philosophical discussion of what has been called the problem of personal identity is generally considered to be concerned with the question of reidentification. Those involved in this discussion say virtually nothing about the self-knowledge question. Instead, they attempt to spell out the necessary and sufficient conditions for saying that a person at time t_1 is the same person as a person at time t_2—to give criteria of personal identity over time which would enable us to answer the amnesiac's question.

The primary contenders for a criterion of personal identity have been the bodily criterion and the psychological criterion, which are based, respectively, on the intuitions that it is sameness of body and sameness of personality which are responsible for sameness of person. Of these two, the psychological criterion has been by far the more widely accepted, and current philosophical discussion of personal identity has focused almost exclusively on attempts to refine and defend this criterion.

In what follows, I shall argue that psychological continuity does not and cannot provide the sort of criterion of personal identity which identity theorists wish to provide. My claim will be that such criteria are inherently circular—that they cannot answer the circularity objection that is standardly raised against them in the literature, and hence cannot provide an analysis of our concept of the persistence of a person.

I shall start, in section I, by giving an exposition of the central features of these psychological-continuity theories. To do this, I shall look in detail at Derek Parfit's[1] version of this theory, and shall use

* At all stages of this paper, I was greatly helped by discussion with friends and colleagues. I would like especially to thank Stanley Cavell, Juliet Floyd, Hannah Ginsborg, Dorothy Grover, Anil Gupta, Nick Pappas, Tim Scanlon, Miriam Solomon, Paul Teller, Jennifer Whiting, and the editors of the JOURNAL.

[1] *Reasons and Persons* (New York: Oxford, 1984); page references in the text are to this book.

0022-362X/90/8702/71–92

© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Parfit as a representative of psychological-continuity theorists throughout the paper. I choose Parfit for three reasons. First, his view is, for my purposes, perfectly representative. There is no argument I make against Parfit which could not be applied to any other standard psychological-continuity theory without significant alteration. Second, Parfit's view is one of the strongest versions of this theory, and, finally, it is the version of the psychological-continuity theory which has been most discussed in the recent literature.

After laying out Parfit's view in section I, I proceed, in section II, to describe the circularity objection that has standardly been raised against such views, and the standard response as offered by Parfit. In section II, I shall argue that this response relies implicitly on a highly implausible view of human experience, and that when this view is made explicit it becomes clear that not only does the standard response fail to overcome the circularity objection, but that this objection cannot be overcome, that psychological accounts of identity, to be accurate, must be circular. Finally, in section IV, I shall suggest that the dead end that psychological-continuity theorists have encountered is a result of their conflating the two questions of personal identity outlined above—the reidentification and self-knowledge questions—and suggest what I believe to be a more fruitful direction for philosophical work on personal identity to take.

I

Parfit begins the statement of his version of the psychological-continuity theory with some preliminary definitions. The first is of *direct psychological connections,* which are connections of the following sort: that between a memory and the experience of which it is a memory, that between an intention and the later act in which it is carried out, and the persistence of a belief, desire, or other psychological feature (205-6). Parfit goes on to define a second relation: "*Psychological connectedness* is the holding of particular direct psychological connections." He then says that "we can claim that there is enough connectedness [for personal identity] if the number of connections over any day, is *at least half* the number of direct connections that hold, over every day, in the lives of nearly every actual person. When there are enough direct connections, there is what I call *strong* connectedness" (206). He thus takes as the criterion of personal identity over time,

> *The Psychological Criterion:* (1) There is *psychological continuity* if and only if there are overlapping chains of strong connectedness. X today is one and the same person as Y at some past time if and only if (2) X is psychologically continuous with Y, (3) this continuity has the right kind of cause, and (4) there does not exist a different person who is also

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

psychologically continuous with Y. (5) Personal identity over time just consists in the holding of facts like (2) to (4) (207).

It is worth considering just what information this criterion is supposed to provide.

Identity theorists take it for granted that we can, in ordinary circumstances, make accurate judgments of personal identity. What they wish to determine is what it is which underlies these judgments. Parfit says that, by 'the criterion of personal identity over time', he means "*what this identity necessarily involves, or consists in*" (203). These theorists are trying to give us an analysis of our concept of personal identity, and so the criterion that they provide will not only have to capture our intuitive notion of personal identity, but it will have to do so in terms more basic. Parfit acknowledges this when he argues that the physical and psychological criteria as he has defined them,

> . . . are both *Reductionist*. They are both Reductionist because they claim
>
> (1) that the fact of a person's identity over time just consists in the holding of certain more particular facts,
>
> and
>
> (2) that these facts can be described without either presupposing the identity of this person, or explicitly claiming that the experiences in this person's life are had by this person, or even explicitly claiming that this person exists. These facts can be described in an *impersonal* way (210).

Parfit is attempting to give an identity criterion that does not employ facts about persons or their identities in its specification, because if such facts were included then this criterion would not provide an analysis of the appropriate sort. It is in the context of this understanding of the project of giving an identity criterion that the circularity objection that I shall discuss in the next section has its force.

II

The circularity objection as it is standardly raised is an adaptation of an objection that was originally raised against John Locke by Bishop Butler. Butler's objection is roughly as follows: memory connectedness seems like a plausible criterion of personal identity, because, properly speaking, we can remember only our own experiences. We can *seem* to remember experiences which are not ours, however, and such seeming memories are no basis for claims of personal identity. A madman may think he remembers leading the troops at Waterloo, but this does not make him Napoleon. Therefore, if memory connectedness is to be a criterion of personal identity, we must have

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

some way of distinguishing between real and merely apparent memory. But that distinction, the argument continues, is no more than this: real memories are apparent memories in which the person remembering is the person who actually had the experience. The obvious problem with the memory criterion of personal identity, then, is that one must already have a criterion of personal identity in order to define memory. Butler argues that, since the fact of identity is prior to the distinction between real and apparent memory, personal identity cannot be defined in terms of memory connectedness.

Psychological-continuity theorists consider a version of this objection, which addresses more general psychological accounts of identity. Parfit indicates that he takes Butler's objection to apply to his psychological-continuity theory when he says:

> On one interpretation, the objection would be this: 'It is part of our concept of memory that we can remember only *our own* experiences. The continuity of memory therefore presupposes personal identity. The same is therefore true of your Relation R [non-branching psychological continuity]. You claim that personal identity just consists in the holding of Relation R. This must be false if Relation R itself presupposes personal identity' (220).

To defend his psychological criterion, then, Parfit needs to show that none of the connections that constitute psychological continuity—memory connections, the connections between an intention and the action that carries it out, or between the temporal parts of a persistent psychological feature—presupposes facts about personal identity.

Before looking at Parfit's response to this objection, it is worth considering why these other connections might be taken to presuppose personal identity. The objection holds that delusional memories are distinguished from real memories by the nonidentity of the remember with the person who had the remembered experience, and thus that when we talk about a genuine memory connection we have already presupposed the persistence of a single individual. The case is similar with the connection between intentions and the actions in which they are realized.

It seems to be part of our conception of intention that I can only properly be said to intend actions that I can reasonably believe it is in my power to undertake. If I am sane and rational, I cannot, in ordinary circumstances, intend to leap tall buildings in a single bound, though I may fantasize about doing so. Similarly, I cannot intend the actions of other people, because they are not under my control. An intention, then, is necessarily an intention that *I* do

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

something, and so it seems that an action will count as the action that carries out an intention only if the same person who forms the intention takes the action. In this way, the connection between an intention and the action in which it is realized seems to presuppose the persistence of a single person.

We do sometimes speak as if we could frame intentions for others. I may, having asked someone to bring me a book, and finding that she has brought the wrong book, say something like, 'That's not the one I intended for you to bring'. Here, I think, this phrase is an abbreviation for 'That's not the one I intended to direct you to bring', and so the intention in question is still an intention that *I* do something. Even if we do not take such phrases to be abbreviations, however, and do admit that we can form intentions for others, it is clear that the only connections between intentions and actions which are going to be relevant to identity over time are those in which I intend that *I* do something, and *I* carry the intention out, and so the connections that concern us between intentions and the actions that carry them out seem to presuppose the persistence of a single person.

The connections between the different temporal parts of a persistent belief, desire, or other psychological feature seem to presuppose identity in a slightly different way. The problem here seems to be that our saying that we have the same belief or desire at two different times requires not only that there are occurrent beliefs or desires at these two times with the same content, but also that these two occurrent beliefs or desires be had by the same person. Countless people can have beliefs and desires that could be described as having roughly the same content, but this connection does not seem relevant to questions of personal identity. The fact that I believe we should avoid an all-out nuclear war at all costs makes me, in this respect, more *like* all the other people who hold that belief, but it seems irrelevant to the question of identifying me with them. Certainly the fact that someone now holds this belief and someone in the future holds this belief does not, in the absence of any further information, give me any information as to whether or not these two people should be identified. My continuing belief that we should avoid nuclear war is not, therefore, relevant to my persistence through time only because it guarantees that my later self will hold beliefs *like* my earlier self, but because the claim that the *same* belief has persisted presupposes that I have continued to exist.

Butler's objection, then, can be expanded to an objection against psychological-continuity theories. In the case of each psychological connection, qualitative identity of the contents of consciousness is

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

not enough to capture what is relevant to personal identity. It is not enough that the madman's memory is qualitatively like Napoleon's experience, that an action taken by someone else can be described as qualitatively like the action I intended, or that my fellow peace lover's belief that we should avoid nuclear war is qualitatively like mine. Something more must be added to this qualitative similarity to make genuine psychological connections. This something more seems to be persistence of a single individual through the two temporal parts of the connection, and this makes relation *R* look circular.

Parfit's response to this objection derives from a commonly used strategy in the contemporary literature. To give a noncircular identity criterion he must give an account of the distinction between genuine and merely apparent psychological connections which does not appeal to facts about personal identity. Like other psychological-continuity theorists, he attempts to make out this distinction in terms of the cause of the later state. He starts with memory, and says of Butler's objection:

> To answer this objection, we can define a wider concept, *quasi-memory*. I have an accurate quasi-memory of a past experience if
>
> (1) I seem to remember having an experience,
> (2) *someone* did have this experience,
> and
> (3) my apparent memory is causally dependent, in the right kind of way, on that past experience (220).

This notion, he says, will answer the objection by giving us a way to distinguish between the kind of apparent memory that is relevant to personal identity and the kind that is not, without appealing to facts about the persistence of persons. Parfit does not explicitly define quasi desires, quasi beliefs, or quasi intentions, but he does recognize their importance within his view, and suggests that they can be specified in the same way that quasi memory is (226).

In the next section, I shall argue that, though quasi states seem initially like a plausible solution to the circularity objection, on deeper reflection it is by no means clear exactly how we are actually to imagine such states, and that on any understanding these states cannot do the work they are supposed to without implicitly presupposing an unacceptably superficial view of the nature of experience. Before turning to my argument, though, it is necessary to consider in a little more depth just how quasi states are supposed to help defend psychological-continuity theories, and to clarify some of the terminology that will be used in what follows. Here I shall discuss mostly

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

memory connections, but anything I say can be extended to the other connections that make up psychological continuity as well.

The circularity objection is based on the claim that it is part of the definition of memory that we can only properly be said to remember our own experiences. In acknowledging the need to respond to this objection, identity theorists implicitly accept this definition of memory, and, in offering quasi memory as a response to the objection, they implicitly claim that it is not this fact about memory which makes memory connectedness a plausible criterion of personal identity. If, as they claim, quasi-memory connectedness will do just as well as genuine memory in specifying an identity criterion, then it must be the case that whatever it is about memory which makes it a central part of what constitutes personal identity is also present in quasi memory. It follows, then, that it cannot be the fact that memory presupposes the identity of the rememberer with the person who had the experience that underlies the crucial role that memory is taken to play in the constitution of personal identity.

Part of what is behind this conviction, I think, is a conflation of what I take to be two separate issues: the nondelusionality of a memory, and its relevance to the constitution of personal identity. The circularity objection itself relies on a claim that memory, by definition, presupposes personal identity. The examples that are given to support this claim, however, rely on the need to use personal identity to distinguish between delusional and nondelusional memories (e.g., between the madman's apparent memory of Waterloo and Napoleon's), and this distinction is a slightly different one.

In defending the use of quasi memory, psychological-continuity theorists present arguments to show that such connections can provide us with a way of capturing the nondelusionality of a genuine memory without the presupposition of personal identity. I shall argue that this does not yet show us that quasi memories capture what is relevant to personal identity in genuine memory, and that this cannot be captured without the presupposition of personal identity. Before making these arguments, though, it is important to have a clear view of how it is that quasi memories are thought to be nondelusional.

The idea, I believe, is this. What makes the madman's apparent memory of Waterloo delusional is not simply the fact that an apparent memory with this qualitative content occurs in his psyche, but rather the fact that he takes this apparent memory to be of an experience that he had. To just have an apparent memory with this content and remain agnostic about whose experience is being remembered would not be to have a delusion on this view, even if the quasi

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

rememberer has had no experience that corresponds in content to the quasi memory. The distinction that is being made, then, can be described as follows: to have a *memory,* is to have an apparent memory of an experience that one actually had, and to take it (correctly) to be one's own experience. To have a *delusion* is to have an apparent memory of an experience that one did not in fact have, and to take it (incorrectly) to be one's own. To have a *quasi memory* is to have an apparent memory (properly caused) and to hold no view about whose memory it is.[2] Insofar, then, as this last case describes a nondelusional apparent memory, quasi memory seems to give us a way of specifying nondelusionality without reference to sameness of person.

That this is the kind of categorization that identity theorists have in mind can be seen in the example of quasi memory that Parfit gives. After giving his definition of quasi memory he says: "We do not quasi-remember other people's past experiences. But we might begin to do so. . . . Suppose that . . . neuro-surgeons develop ways to create in one brain a copy of a memory-trace in another brain. This might enable us to quasi-remember other people's past experiences" (220). He then goes on to offer an example, the story of Jane and Paul, in order to show that it is possible for one person to have another person's memory nondelusionally. We are to imagine that Jane has copies of some of Paul's memory traces implanted in her brain, and shortly thereafter has a new vivid set of apparent memories of experiences that she knows she never had. In particular, she has an extremely vivid memory of a storm in Venice and of a bolt of lightning hitting the bell tower of San Giorgio, though she has never been to Venice. On asking Paul, she discovers that he did have exactly such an experience. Parfit concludes: "Given all of this, Jane should not dismiss her apparent memory as a delusion. She ought to conclude that she has an accurate quasi-memory of how this flash of lightning looked to Paul" (221). Jane, then, has an accurate recreation of the qualitative content of Paul's memory of seeing the flash of lightning, and it is (by Parfit's definition), appropriately caused. This qualitative similarity is to be taken quite seriously; we are really to think of Jane as having Paul's memory exactly recreated, and, in regard to this one incident, experiencing exactly the inner state Paul would in a genuine memory.

Having set up the example, Parfit immediately considers the objection that for Jane to be given Paul's memory in this way would

[2] I am extremely grateful to Paul Teller both for suggesting the necessity of laying out this distinction, and for his helpful comments in formulating it.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

amount to her having a delusion, and he dismisses it. He says: "It may be claimed: 'Since Jane seems to remember *seeing* the lightning, she seems to remember *herself* seeing the lightning. Her apparent memory may tell her accurately what Paul's experience was like, but it tells her, falsely, that it was *she* who had this experience' " (221). If this objection stands, then quasi memory will be no help against the circularity objection. But Parfit has a response. He says:

> Because we do not have quasi-memories of other people's past experiences, our apparent memories do not merely come to us in the first-person mode. They come with a belief that, unless they are delusions, they are about our own experiences. But in the case of experience-memories, this is a separable belief. If like Jane we had quasi-memories of other people's past experiences, these apparent memories would cease to be automatically combined with this belief (222).

This, then, supports my claim that quasi memory is thought to avoid the problems of the circularity objection because it can characterize a nondelusional memory without reference to personal identity.

In the remainder of this paper, I shall argue that quasi memory fails to do the work it is supposed to, avoiding the circularity objection because relations between the specification of the qualitative content of a memory, presuppositions about the identity of the rememberer, and the issue of nondelusionality are far more complex than identity theorists recognize. I shall argue that simply deleting the "nametag" from a memory is not sufficient to make it nondelusional, and that in order to make an apparent memory truly nondelusional one will either have to presuppose the identity of the rememberer with the person who had the experience, or else remove so much of the content of the memory that it is no longer plausible to say that what is relevant to personal identity in genuine memory is preserved in quasi memory. So I shall argue that not only do quasi states not fulfil the role that they are supposed to, but furthermore that they cannot—that there is no way to capture what is relevant to personal identity in memories without presupposing identity.

III

I shall start my argument with a detailed look at an example offered by Edward Casey in his book, *Remembering: A Phenomenological Study*,[3] where he describes one of his own memories. If we ask what would happen if this memory experience were given as a quasi memory, the limitations of quasi states will become clearer. Casey recalls a memory of a recent family outing:

[3] Bloomington: Indiana UP, 1987.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

I recall going to the movie *Small Change* a few weeks ago—exactly when, I am not certain. After dinner nearby at Clark's, my two young children, my wife, and I had walked briskly over to the Lincoln Theater, stopping briefly at a paperback bookstore on the way. Anticipating a large crowd, we arrived early and were among the first to purchase tickets. There ensued a wait that seemed much longer than the ten or fifteen minutes it actually was. The children were especially restive and had difficulty staying in the line that had formed—Erin attempting some gymnastic tricks on the guardrail by the entrance, Eric looking at the posted list of coming attractions. Finally the doors were flung open, and we entered at the head of what was, by then, a considerable line. Once inside, we sought seats approximately in the middle of the theater, settled there, and interchanged positions a couple of times to adjust to the height of those sitting in front of us. The lights dimmed, and *Small Change* began directly. (Or was there not a short feature first?—I cannot say for sure.) The film was in French, with English subtitles. I have only a vague recollection of the spoken words; in fact, I cannot remember any single word or phrase, though I certainly remember the characters *as speaking*. The same indefiniteness applies to the subtitles, at which I furtively glanced when unable to follow the French. Of the music in the film I have no memory at all—indeed, not just of *what* it was but *whether* there was any music at all. In contrast with this, I retain a very vivid visual image of the opening scene, in which a stream of school children are viewed rushing home, seemingly in a downhill direction all the way. Two other scenes also stand out in my present recollection: an infant's fall from the window of a high-rise apartment (the twenty-ninth floor?) and the male teacher (whose name, along with all others in the film, I have forgotten) lecturing passionately to his class about child-abuse. Interspersed between these scenes is a medley of less vividly recalled episodes, ranging from fairly distinct (the actions of a child-abusing mother) to quite indistinct (e.g., children's recitations in the classroom). While I am recollecting this uneven and incomplete sequence of filmic incidents, I find myself at the same time remembering my own children's ongoing reactions to the film. I do not remember their behavior in detail but only as a kind of generalized response consisting of laughing, whispered questions, outright comments, and the like. These reactions are as intrinsic to the memory as is the unfolding of the film itself; so too is the mixture of pleasure and exasperation which I felt in being located, as it were, *between* children and film. Suddenly my memory of *Small Change* comes to an end: the lights go up, and we leave through a side exit near us, overhearing expressions of amusement and satisfaction from those around us as we walk out into the night (*ibid.*, pp. 25–6).

Parfit's claim is that our relation to the experiences we remember (taking them to be our own) is separable from their content, and that, if this relation is changed, we can be given someone else's memory

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

nondelusionally. There are a number of features of Casey's memory which show this view to be problematic.

One striking difference between this example of an actual memory and the fictitious example that Parfit gives us is the detail and richness of the former as opposed to the latter. Casey's memory of the trip to see *Small Change* includes memories of his family, the restaurant where he has eaten more than once, the theater at which he has seen more than one movie, and so on. Certainly memories that consist of a striking visual image, like the memory of Venice that Parfit describes, are one sort of experience memory, but more of our memories are like the one Casey describes; they are memories of family vacations, or writing a paper, or turning thirty. These memories are much more complex due to the number of personal associations and details about the individual's life which are involved in them. Intuitively, it seems that these memories are more likely to be the sort that lend plausibility to the idea that memory connectedness is an essential feature of personal identity than memories of the sort Parfit describes, and in any event, since quasi memory is meant to provide a defense of psychological-continuity theories, memories of this richer sort will also have to be possible quasi memories.

This fact, however, commits Parfit and other identity theorists to some quite implausible views about the inner lives of persons. To see this, let us consider what it would be like for Jane to be given Casey's memory of *Small Change* as a quasi memory. The amount of personal detail that this memory includes makes it difficult to imagine Jane receiving it as a quasi memory. Some striking visual image is an experience that could, given the appropriate circumstances (i.e., if Jane had really been in that spot in Venice at that time), be had by virtually any sighted person, regardless of the rest of her psychological makeup. Casey's memory is not like that; it contains a good many elements that make reference to other parts of his life and his personality. It includes, for instance, the familiarity with the town in which he lives, the restaurant where he eats, the theater where he has seen numerous movies, and so on. His knowledge or lack of knowledge of French also plays a role. More important, perhaps, is the fact that his relationship to his children is central in what he remembers; his response to the gymnastics on the guardrail, and his mixture of pleasure and exasperation at their responses to the movie are, he says, essential parts of this memory.

These features of the memory are going to be very alien to Jane, who presumably does not have a wife, and, we may assume to make our point, does not have children. Not only will the physical locations seem unfamiliar, but so much else about the memory will seem

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

anomalous. If Jane speaks fluent French, then the experience of reading the subtitles will be puzzling. And if she happens to have acted in this movie, or to be a musician, then her failure to recognize herself, or to remember whether there was music will be very disturbing. Furthermore, the idea of having a wife and children, as well as the interest/pride/embarrassment/concern relating to the behavior of these particular children, remembered as Erin and Eric, will be out of place in Jane's psyche.

Once we have noticed that this memory contains so many elements personal to Casey and anomalous to Jane, it is no longer easy to imagine what her quasi memory will be like. It seems to me that there are only two possibilities. The first is that Jane will reproduce all of the visual content of the memory without interpreting it as Casey does. That is, upon awakening from the quasi-memory implant surgery, Jane will have images of being in a strange restaurant, then a bookstore, then a theater, with a woman and two children whom she does not recognize, and she will also have images of seeing a movie with these people. The second alternative is that she will reproduce the memory exactly as it occurs in Casey, with all of the same personal elements and associations. Neither of these alternatives will allow quasi memory to do the work it is supposed to.

Consider the first alternative. It is, first of all, not entirely clear that this alternative is even actually possible. Our memories may not be strictly visual in the way that this alternative suggests. We should not take it for granted that Casey's remembering his family, for instance, is his having visual images of some people whom he is able to recognize. Instead he may remember them nonvisually as his wife and children, and this would be impossible to reproduce neutrally in Jane. Even if there were a way to reproduce this memory in some neutral way in Jane's consciousness, though, this alternative would fail because quasi memory of this sort would not capture what is relevant to personal identity in genuine memory connections. If so much of what this memory is to Casey is missing, if instead of a memory of a family evening out there is only a series of images, then we have little basis to say that Jane's quasi memory is qualitatively the same as Casey's memory, even though it does contain many of the same elements. Since memory is serving as a possible basis for personal identity, it seems that the personal elements, the ones that will be missing in Jane, are likely to be the most crucial.

There is good reason to question whether Jane's experience, so understood, is qualitatively like a memory at all. Phenomenologically, it will be drastically different from Casey's experience, appearing not as a coherent memory of a family outing, but as a blur of

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

unidentifiable sights and sounds, which will make it not only unlike Casey's memory, but unlike any memory. If this quasi memory is different from Casey's experience in such a fundamental way, then we seem to have no good reason to call it a quasi memory of Casey's experience. This in turn makes it seem quite unreasonable to say that this quasi memory captures what is relevant in the connection between a genuine memory and the experience remembered.

The second possible way to imagine this case is to imagine Casey's memory reproduced in Jane exactly, with all of its personal elements. In this quasi memory, then, it will seem to Jane that these people are her wife and children; Clark's, the bookstore, and the Lincoln Theater will seem familiar, and the emotional responses to the movie and to Casey's family will seem like her own. If this is what happens, though, then it is difficult to make out Parfit's claim that the belief that an experience I remember is my experience is separable from the memory itself. If the memory must be such that I think of it as *my* family and *my* hometown, then the mineness of the experience seems to be part of the content of the memory. Parfit might respond that, whereas that is true to a point, the belief is separable at a later stage; whereas the apparent memory itself may carry the semblance of the experience being hers, on reflection Jane would be able to separate this out. This would involve her having a memory qualitatively just like Casey's, but reflecting on it and saying to herself something like, 'These people seem to be my family, so they must be the family of whoever's memory I've been given, for they certainly aren't mine'.

There are problems with this solution, too, however. The first is that it is again not clear that we are entitled to say that Jane's quasi memory is qualitatively the same as Casey's memory. Although it contains all of the same elements, there are additional elements that will be added by Jane's psyche. If Jane has no children, no spouse, never goes to movies, and has never been in Casey's hometown, then surprise and confusion are likely to accompany this quasi memory in Jane. It will seem extremely foreign, anomalous, and as if it cannot be right but must be some sort of delusion or fantasy. This is not just a memory that Jane cannot quite place; it is one that cannot, in principle, be a part of a coherent life history for her, since it contradicts what she knows about herself. A memory that is familiar and nostalgic to Casey seems foreign and fantastic to Jane. It seems quite plausible, therefore, to say that this memory, with its dimension of strangeness, is a qualitatively different experience for Jane than for Casey. If, therefore, we really wanted to reproduce the qualitative content of Casey's memory in Jane, we would not only have to recreate a great many of Casey's states in Jane, but suppress a great

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

many of Jane's as well, and this begins to look suspiciously like replacing Jane's psychology with Casey's.

Even if we took this route, though, Parfit would still be in trouble because Jane's memory would no longer be nondelusional. We have seen that giving Jane the full content of Casey's memory requires the inclusion of facts about who has the memory. These facts, however, are false in relation to Jane, so their inclusion will make her quasi memory delusional. For Jane to experience something like Casey's memory, she will have to take his family to be her family, his town to be her town, his emotional reactions to be her emotional reactions, and so on. But these are not her family, town, and emotional reactions any more than Napoleon's troops are the madman's, and as long as Jane has to take these components of Casey's memory to be her own, there is no way to make this memory nondelusional.

This discussion has shown that quasi memory cannot do what it was devised to do, because presuppositions about who has a psychological state come in at a level deeper than the level of the connections between states; they are necessary to defining those states as well. The original circularity objection charged that specifying the connections between psychological states required appeal to facts about the identity of the person who had the states. Parfit's original response, therefore, was that he needed to find an impersonal way of differentiating between two qualitatively identical apparent states (say, the madman's memory of Waterloo and Napoleon's), one of which was genuinely connected to an earlier state and one of which was not. It was presumed that the content of the memory could be defined independently of facts about whose memory it was, and that all that was needed to avoid circularity was to show that the relevant connection of this impersonally defined memory to an impersonally defined earlier experience could also be defined impersonally. This is what Jane's case was supposed to show.

What I have argued, though, is that facts about whose memory a given memory is are an integral part of its qualitative content. Without the associations to facts about the rememberer's life and psychology which a memory ordinarily has, an experience loses so much of its content that it seems arbitrary to call it even the same apparent memory, and so to take it to be in any way relevant to questions about personal identity. If, on the other hand, the entire content of the apparent memory is reproduced, then the inclusion of the facts that allow this will guarantee that the memory is delusional. The fact, then, that presuppositions about who has a memory are inseparable from its content means that one cannot, as Parfit claims, specify

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

nondelusionality impersonally by keeping the content of a memory and simply deleting presuppositions about whose memory it is.

The case is no different for the other quasi states, which encounter the same difficulties as quasi memory. Suppose, for instance, that instead of a memory of Venice it is the desire to join the French Foreign Legion which is transferred from Paul to Jane. How are we to imagine this? If it consists in no more than Jane suddenly having the thought 'I'd like to join the Foreign Legion', then it is not at all clear why we should even call this Paul's apparent desire, or think that its occurrence in Jane makes her any more psychologically continuous with Paul than the fact that they both like chocolate ice cream. If, on the other hand, this desire occurs in Jane with all the elements that make it Paul's desire—the belief that a soldier's life is romantic and exciting, the boredom with his job as an accountant, the chaos in his personal relationships which makes him want to run off to a foreign country, and his childhood fascination with *Beau Geste*—then it is clear that we are imagining Jane's changing in ways far more fundamental than Parfit indicates. It is easy to imagine examples that illustrate that quasi beliefs, quasi values, and quasi-character traits face exactly the same problems.

Parfit faces the same dilemma with respect to these states as he does with respect to memory. If we imagine Jane's being given a desire that is like Paul's only in the bare minimum of content, then this similarity is too weak to provide a genuine psychological connection between Jane and Paul. Quasi desires were defined, I said in section II, precisely to differentiate between desires that correspond merely in this minimal sense and those which are truly persistent. If we are to be given any reason to consider this desire as one that starts in Paul and persists in Jane, then it will have to at least contain some of the features that make it uniquely Paul's, and that means including his memories and associations.

If these are included, however, then the desire will fail to satisfy the requirement that it not presuppose facts about who has it, and it will be delusional. The chaos in Paul's personal relationships and the boredom with his job would have to be the chaos in *his* relationships and the boredom with *his* job. They would have, furthermore, to appear in Jane as chaos in *her* personal relationships and boredom with *her* job if she is to have a desire qualitatively like Paul's and this would make such a quasi desire delusional.

This point is even clearer in the case of what have been called in the literature *personal desires*. These are desires which explicitly include as part of their content that the person having the desire be

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the one to fulfil it. For example, I may have a desire to win the tennis match that I am playing this afternoon. This desire is specifically the desire that *I* win the tennis match. My opponent probably has a desire that overlaps minimally in content, and that is his desire that *he* win the tennis match. Both of these have content that could be described from the first person point of view by 'I want to win this match', but they are certainly not the same desire, and the fact that they overlap in content in this way certainly does not make me psychologically connected with my opponent with respect to this desire.

Now try to imagine what it would be like for Jane to be given this desire as a quasi desire. If she desires that *I* win the tennis match, then her experience will be different from mine because it will not include the anticipation of the play of the game, the feeling that she, the desirer, had been beaten by my opponent one too many times, or the anticipation of standing in front of the spectators exalted if I win and humiliated if I lose. We all know that it is qualitatively different to desire something for someone else than to desire it for ourselves, even if the desire is just as strong in both cases. If, on the other hand, Jane desires that *she* win the tennis match, then her desire will be qualitatively different from mine in just the same way that my opponent's is, unless she takes herself to be me, in which case she would be deluded. Quasi states, such as belief and desire, fail to fulfil their function for just the same reason that quasi memory does, because the content of these psychological states, too, cannot be defined without presuppositions about who it is who has them.

What this discussion has shown is that, on either of the two possible pictures of what it is to have a quasi state, quasi states fail to do the work they are supposed to because they include either too little or too much of the state they reproduce. If they include too little, they do not capture what is relevant to personal identity, and if they include too much, then, unless sameness of person is assumed, they are delusional.

IV

My objection is not just an objection against quasi states, but also against the project that includes them. The attempt to give a noncircular analysis of identity over time requires that the distinction between genuine and apparent continuity be made without presuppositions about the identity or existence of the person having the states which constitute such continuity. This turns out to be impossible. My argument that quasi states cannot serve this function has not relied on any particular facts about such states as Parfit defines them; it has relied only on the fact that we cannot imagine a state qualitatively like

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Paul's occurring in Jane without that state being delusional. The mineness of a psychological state cannot be separated from its content, and so, to define a state that can be properly called one of the states constitutive of personal identity, we will have to presuppose the existence of a persistent person who is the subject of that state. This means that the strong noncircularity condition cannot be met, and so an analysis of personal identity of the type identity theorists wish to give is not possible.

It would certainly be unfortunate if this meant that we had to give up on psychological accounts of personal identity altogether, since the intuition that psychology and personality are essential to our identities is extremely strong. Luckily, the above discussion does not have this consequence. It does, however, require us to rethink what such criteria will be like, and what will be required of them. I would like to conclude, then, by offering a sketch of an alternative way of conceiving of what it is to give a psychological account of identity. This alternative promises to capture many of the intuitions that support Parfit's psychological-continuity theory without running into the problems which his view does. Such a view will not provide us with a reductionist analysis of personal identity of the sort that psychological-continuity theorists try to offer, but it can provide us with a great deal of insight into what is involved in the identities of persons by emphasizing rather than neglecting the complexities of our psyches.

In order to understand this alternative better, it is helpful to return to the issues that I raised at the beginning of this paper. There I suggested that the problems that psychological-continuity theorists encountered could be explained by their conflation of the two questions of personal identity: the question of reidentification, on the one hand, and of self-knowledge, on the other. I cannot fully argue for this claim here, but I can say something about how such an argument would proceed, and this will be helpful in developing an alternative psychological account of personal identity.

It is a philosophical commonplace that we have a dual perspective on persons. On the one hand, we view persons as one of the types of objects in the world, but, on the other, we view them as subjects and agents, creatures with a way of experiencing the world and with affect and volition. Because we have these two views, we tend to use the word 'person' in two different ways. Viewing persons as objects of our knowledge, we tend to use the term as if it were more or less coextensive with 'live human body'. Viewing persons as subjects and agents, we tend to use 'person' as an honorific term, having to do

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

 Digitized by Arya Samaj Foundation Chennai and eGangotri

with autonomy, moral agency, and volition. In this sense, not all human beings will necessarily be persons, and not all actions of human bodies will be actions of persons.

It is my claim that the two questions of personal identity come from these two perspectives also. Insofar as we view persons as objects of our knowledge, it will make sense to view the question of personal identity as a subspecies of more general metaphysical questions about the identity of complex objects, and to employ the same methods that are used in this pursuit to answer questions of personal identity. These methods involve looking at time slices of persons from an atemporal third-person perspective, and attempting to give objective criteria for saying that two time-slices belong to the same person. Ultimately, I believe, these methods can only support a bodily criterion of personal identity. Viewing persons as subjects and moral agents, on the other hand, discourages such an objective approach to identity, and leads us to emphasize the subjective experience of a person and questions of responsibility in addressing questions of identity (the sorts of identity questions raised in ethics, action theory, and literature, for example).

My claim is that contemporary psychological-continuity theorists like Parfit have taken their goal of providing a noncircular identity criterion from the reidentification question, and the intuitions that support the psychological criterion over the bodily criterion from the self-knowledge question. It is this, I claim, which has pushed these theorists to hold such an implausible view of our psychological states. The arguments that support a psychological criterion all rely on hypothetical puzzle cases, in which we are asked to imagine someone changed in certain ways, and then to observe our intuitions about whether or not the original person should be concerned about the future of the resultant person, held responsible for her actions, or take her beliefs to be her own. Issues of agency and self-knowledge are emphasized, and it is on the basis of these that we are being asked to judge questions of identity. The assumption is that when we concentrate on such questions we will judge that responsibility and concern (and hence identity) go with the psychology or consciousness rather than the body.

These thought experiments, then, are taken to support the view that psychology rather than the body is the constitutive factor of identity, but the question is never raised as to what kind of thing it is the identity of which is at issue—the person taken as subject or as object. The intuitions we have in response to the thought experiments, I claim, come from our view of persons as subjects, but the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

methods that identity theorists use to turn these intuitions into a criterion of identity come from a view of persons as objects. Taking the fact that psychology is what turns out to be important in these cases, psychological-continuity theorists thus make the unwarranted assumption that sameness of psychology can be used to provide a noncircular criterion of identity of the sort which is given for objects. But such a criterion cannot focus on subjectivity; it is, by definition, to be objective, and must be capable of being spelled out without including the first-person perspective of a given individual. The pieces that make up a person's psychology, must, to fulfil this purpose, be viewed to be as discrete and detachable as are the planks of a ship or the grains of sand in a heap. It is because psychological-continuity theorists are trying to force the insights gained from consideration of questions of self-knowledge and responsibility into the mold of questions of the persistence of material objects that they are forced to view psychological states as atomic, isolable, and in principle independent of the subject who experiences them—a view that I have argued to be highly implausible.

The intuitions that support psychological criteria come from the considerations that are at play in the self-knowledge question, and so it seems reasonable to hope that the sort of work done to address this question will be of some help in reconceiving the notion of a psychological account of personal identity. Let us consider this question for a moment. The person interested in self-knowledge will ask herself questions of the following sort: 'Is this really what I believe?' 'Is this really what I desire?' 'Is this really what I intend?' Which is to ask, in a specific way, 'Is this my belief?' 'Is this my desire?' 'Is this my intention?' The resources involved in answering these questions would be resources that allowed a person to assign particular psychological states to herself, and so would be at the same time resources for delimiting an individual person, and for answering questions of personal identity.

What kind of information, then, can be used in answering these questions? We have seen one relevant set of considerations already in the discussion of quasi states. In each of the cases we imagined, we could identify something highly anomalous about the states given to Jane even independently of the knowledge that they had come from someone else. We saw that Casey's memory or Paul's desire to join the French Foreign Legion, put into Jane's psyche in full form, were not only delusional, but jarring. This is because they contained so many details that were in conflict with the details of Jane's own life. Jane, I claimed, would be unable to appropriate these states coher-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ently as her own, because they would conflict so violently with what she knows to be true of herself. If Paul's desire to join the French Foreign Legion were, therefore, to occur in Jane's psyche, she would have good reason to doubt that it was really her desire, even if she had no prior knowledge of its origins.

Part of what is involved in a psychological state's being mine in the sense which is at issue in the self-knowledge question, then, is its coherence with my total psychology—my ability to view it as a comprehensible part of my life, and to take it to be my own. Clearly the sort of coherence required here cannot be strict logical consistency —few of us are fortunate enough to have psyches without conflict, ambivalence, or discrepancies in our memories—but there is some way in which we are able comfortably to appropriate the various aspects of these ordinary conflicts, and view them as part of a single, comprehensible life. When the conflict is too severe, however, and the states in question are incomprehensible to us within the context of our self conception—as Casey's memory or Paul's desire might be to Jane—then these states have the character of external impositions, and a question is raised as to whether they are a genuine part of the person in whom they occur.

It is precisely when someone runs into an anomaly of this sort that we say that she is having an "identity-crisis," and the resolution of this crisis generally involves the forming of a well-grounded self-conception, and the ability to reject as not one's own beliefs, desires, and character traits that are not comprehensible as part of that self-conception.[4] As an alternative approach to offering a psychological account of personal identity, I propose that we take this talk of "identity crises" literally and not metaphorically—that we take persons (viewed as subjects and agents) to be constituted by their own self-conceptions. This view makes sense both in light of the discussion of quasi memory—which shows that the inability of someone to appropriate an experience coherently is sufficient reason to say that it is not hers—and in light of the intuitions that support the psycho-

[4] I cannot emphasize enough that this is the sketch of the view, and that many distinctions and definitions are required to fill it out. What I want to make clear here is that what I have just said does not mean, for instance, that, if I conceive of myself as a patient person, and find suddenly that I have murderous impulses concerning the person driving five miles an hour in front of me on a one-lane road, I can simply reject these murderous impulses and thereby make them not mine. I understand exactly where these impulses come from, and am forced, in the face of them, to revise my view of myself as a patient person. The type of incomprehensibility that I have in mind here is stronger and of a different sort. I shall say more about this later in the paper.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

logical criterion in the first place. These intuitions, I argued, are based on a notion of persons as self-conscious agents who are capable of moral responsibility, and such a conception seems to require as an essential feature of persons that they have a coherent self-conception. It is, after all, precisely when we find that persons do not have a coherent conception of themselves as agents that we feel we cannot legally hold them responsible for crimes they have committed—that there is a relevant sense in which these crimes are not their actions.

Furthermore, there is much philosophical precedent for viewing persons as self-constituting, the fact of their identities created by their self-conceptions. This view is central in the work of the existentialists, but it can also be found closer to home in the work, for instance, of Harry Frankfurt,[5] or Herbert Fingarette.[6] It is also very much present in the work of Locke[7] himself, who tells us that a person is "a thinking intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places" (*ibid.*, p. 39), and that "Wherever a man finds what he calls himself there, I think, another may say is the same person" (*ibid.*, p. 50). Since contemporary psychological-continuity theorists take themselves to be addressing the question that Locke raises, it does not seem too far-fetched to believe that a view of persons as created by their self-conceptions, which Locke himself proposed, could provide a means of answering this question.

Of course, this view needs a great deal more development. I have already mentioned the need to spell out in more detail both what a self-conception is and what coherence is. The madman with Napoleonic delusions takes himself to have led the troops at Waterloo, but this does not count toward making that his action. And my refusal to accept my competitive impulses as my own does not have the consequence that I am not a competitive person. Also, it seems clear that I cannot use just *any* material to form my self-conception. Certain things I must view as part of my life if I am ever to have such a conception, and other things I cannot: there is no coherent way for me to appropriate Napoleon's actions at Waterloo. More must be

[5] See, for instance, his "Identification and Externality," in *The Identities of Persons*, Amelie Oksenberg Rorty, ed. (Los Angeles: California UP, 1975).

[6] See, for instance, his *Self-Deception* (London: Routledge & Kegan Paul, 1969).

[7] "Of Identity and Diversity," from *An Essay Concerning Human Understanding*, reprinted in *Personal Identity*, John Perry, ed. (Los Angeles: California UP, 1975).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

said, then, about what it is which determines what can be involved in my self-conception, and about what a self-conception is.

Fortunately, we already have many resources available to help in filling in the gaps which we find in this view. There are, first of all, the philosophical sources that I mentioned above. We can also learn a great deal from work in psychology. Developmental psychology tells us how an infant progresses from a creature who does not distinguish between itself and the external world, to a person who distinguishes clearly between what is himself and what is other. And abnormal psychology provides us with analyses of cases where the attempt to form an identity goes awry. There are also many rich suggestions in literary treatment of identity, metamorphosis, and the double, which can be of tremendous assistance in trying to spell out this view. There is good reason to hope, then, that an alternative psychological account of persons of the sort I propose could be developed.

A view of persons as self-constituting would, when developed, provide an identity criterion of sorts. If we wanted to know, for instance, whether Jane is the person who had some particular experience, then we would need to discover whether this experience is a part of her coherent self-conception, because it is this which is the necessary and sufficient condition for assigning an experience to a particular person. This criterion, it is clear, would not fulfil the strong noncircularity condition to which psychological-continuity theorists hold themselves. It is not a reductionist analysis of persons, because it requires that in each instance we settle questions of identity with reference to the subjectivity and psyche of the person in question. I hope that the sketch I gave above will be sufficient to show, however, that this circularity is nonvicious—that such a criterion can provide us with a great deal of insight into what it is which constitutes the identities of persons, and that it can do so without doing violence to the complexities of our experience.

MARYA SCHECHTMAN
University of Illinois/Chicago

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

BOOK REVIEWS

Neurophilosophy: Toward a Unified Theory of the Mind/Brain. PATRICIA CHURCHLAND. Cambridge: MIT Press, 1988. 546 p. Cloth $27.50.

In the introduction to *Neurophilosophy,* Patricia Churchland states that its central task is to address "the overarching question of the nature of a unified, integrating theory of how, at all levels of description, the brain works"—that is, to provide good reasons why a unified theory of the mind/brain is possible and probable, and to make a start in defining that theory. Churchland's secondary goals concern her audience. Because the central theme of the book is unification, Churchland seeks to address both philosophers and neuroscientists, to educate both groups. On the home front, she hopes to give philosophers a sufficient background in neuroscience to sustain the examples used and to enable philosophers to forage the literature themselves; for the neuroscientists, the goal is to provide a "get-around" knowledge of the philosophical issues.

Significantly, given our materialist goals and tendencies, *Neurophilosophy* is the first book of its kind. It is a vital and timely project —moreover, one to which Churchland's talents are uniquely suited.

Neurophilosophy does not, however, accomplish all that it attempts. The central reason for this failure is not philosophical per se. Rather, given the magnitude of the goals coupled with the division in audience, the scope of this project is too wide—what *Neurophilosophy* does is to compress the work of several volumes into one. Desiring breadth, it has sacrificed detail, the result being a work which is both philosophically sketchy and, at points, unclear in its intent.

That said, this review will begin with a brief discussion of Churchland's theory of the mind/brain. My aim is to say how the agenda for the book is set by the philosophical view—to say, on the theory's own terms, what it would take to complete the project—and then to contrast this with what Churchland has actually written. The second half of the review is a discussion of a major piece of the project, Churchland's theory of content.

I. THE PROJECT

At the foundation of Churchland's view of the mind/brain are two central theses. The first thesis, from the philosophy of science (Pierre Duhem, W. V. Quine, Imre Lakatos), states that the terms of a scientific theory have meaning only within the context of that theory.

0022-362X/90/8702/93–102 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

There are no "pure" observational predicates that take their meaning "from their referents"; there are only terms which are embedded in the theory as a whole, which are "theory-laden." The second thesis, ascribable to Wilfrid Sellars, is about the nature of our understanding of mental events. Here, we are asked to regard the compendium of our knowledge of mental states as a theory—and, moreover, a theory whose hallmark features have been modeled, rightly or wrongly, upon the "visible" or intersubjective features of language, on utterances, not upon the results of introspection. Putting these two theses together yields the following conclusion. Just as there are no pure observational predicates in theories that purport to describe the external world, the inner eye offers no privileged "theory-free" insight about the nature of mental events.[1] Hence, qua theory, folk psychology might well be wrong, revisable in the light of a better theory—in this case, in the light of our future neuroscience.

Starting with these theses, the question of the future of psychology, of the "scientific" and "folk" variety, arises. Churchland's central claim is reductivist. When all is said and done, the phenomena described by scientific psychology will be explained by the functional entities postulated by a future neuroscience. Even though a traditional "reduction" (of the entities postulated by the one theory to the entities of the other) will probably not be forthcoming, an understanding of the micro-level phenomena (neurophysiological events) will explain how the macro-level events (those described by scientific psychology) come about. On the other hand, *vis-à-vis* the properties and entities of our present folk psychology, the new theory may well prove less kind. Here, elimination (as opposed to reduction) is the most plausible outcome, although, as Churchland emphasizes, this issue is purely empirical.

A commitment to the results of science notwithstanding, Churchland has her money firmly placed: she takes the recent data to suggest strongly, if not quite an outline of the new theory, at least an eliminative scenario. Here, we are likely to find that: (a) many folk psychological categories, such as memory and learning, will not be legitimated by the new functional categories of neuroscience (i.e., there will turn out to be many diverse and different mechanisms for phenomena we took to exemplify one psychological category); (b) the form of the mechanisms underlying thought will be in no way "sen-

[1] I should note here that, although Sellars does arrive at this same conclusion, it is not via the same route: Sellars' argument depends upon a theory of thought coupled with a separate theory of sensation. This correction was pointed out to me by Wright Neely and I thank him for it.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tential" (in Jerry Fodor's sense of the word); (c) the folk notion of "consciousness" will not delimit a natural kind; and (d) on the new theory, although it will still be legitimate to speak of the determinate, objective (albeit relational) meaning of a neural state, we will probably not be able to express that content as a propositional attitude—mental content will be best stated in neurophysiological/computational terms. (This is only a partial exposition of Churchland's scenario.)

Given the above project, then, what would be necessary to give it substance? What "books" are required in order to complete Churchland's philosophical and theoretical project?

(1) Reduction and the Unified Theory of Mind. This book gives an exhaustive accounting of the reductive strategy, including a defense of the central theses given above, the details of the envisioned theoretic reduction, and responses to its standard philosophical objections. Here, we want to know the details. What, for example, is the relation of folk theory to scientific theory? Assuming that we do not represent our folk knowledge by means of mental sentences, in what ways is folk theory a theory? And so on. This is philosophy of mind in the traditional mold.

(2) Why Neuroscience Matters. Part of the problem with claiming that we can treat all knowledge of ourselves as empirical is that the conclusion, however well-argued, remains intuitively unconvincing. Examples are needed. It is only by saying how specific folk psychological theories/presuppositions/dogmas about the mind are now in question, or which ones would be, given specific neuroscientific results, that the eliminativist position acquires bite. It is in this book, then, that the doctrines of folk psychology are made explicit and we are shown how the scientific data tell against them.

(3) The Unified Theory. This book represents a "best bet" about the new theory of the mind/brain. Building upon the conclusions of book 2, the elimination and revision of folk concepts, it gives the positive view—the kind of neurophysiological/computational theory that could incorporate the overall framework.

Next, as Churchland suggests, if the project is to reach both neuroscientists and philosophers, there is the problem of audience, of bringing the two groups up to speed. This would require two texts.

(4) Neuroscience for Philosophers: A Guide to the Principles, Techniques, and Current Findings of Neuroscience. For those philosophers who are skeptical of the project at the outset, such a lengthy primer will seem an onerous and unnecessary burden. Nonetheless,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

learning the neurobiology is not merely a matter of gaining the wherewithal to understand specific examples. Rather, if Churchland's major goal is to find a theory of mind that "makes biological sense," it will be difficult to explain, at any critical juncture, the drift of that phrase. 'What makes biological sense' refers to the kinds of solutions evolution has found to informational/environmental problems. And just as grasping the nature of, say, proof theory is best accomplished by doing (or going over) a number of proofs, understanding the principles of neuroscience is best achieved by walking through a number of evolutionary "solutions." Hence, in order to understand the project as a whole, a neuroscience primer is needed.

(5) *Philosophy for Neuroscientists: An Introduction to the Philosophy of Mind and Science.* Unlike neuroscience, where there is an accepted body of introductory "facts," philosophy has no agreed-upon set of best theories. Not that it would be impossible to select a few "very bad ideas," those ideas we love to hate. Still, because most current theories are highly controversial, an introduction to the philosophy of mind and science would include a summary of alternative views, with the author's biases clearly flagged as such.

Let me turn now to *Neurophilosophy*. Part I, approximately the first half of the book, is a neuroscience primer (book 4) supplemented with some hints as to which experimental areas the author expects will prove theoretically interesting. This is a job well-done—a lucid and knowledgeable work.

Part II is an overview of the philosophical issues, "Recent Developments in the Philosophy of Science." In these chapters, Churchland gives a short history of the philosophy of mind, sketches the outlines for a theory of reduction, and takes up some standard objections. There is also a chapter on the antireductionist sentiments of "functionalist" theories or, more properly, sentential/computational views. The central problem here is that Churchland has combined books 1 and 5, the introduction to philosophy with the argumentation for the philosophical project. As one might suspect, this is a less than harmonious union: often the line between explication and argument is not clearly demarcated (of particular trouble to the neuroscientific reader); for the most part, the arguments needed to sustain Churchland's view are given only in skeletal form, referring the reader to previous articles for further detail.

Part III, the shortest section, is designed to cross the divide. Churchland presents three very interesting and suggestive neurophysiologically-based theories—tensor-network theory, the vestibu-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

lar-ocular reflex, and PDP models—each chosen in the hope that it might exemplify "the foundations of a new paradigm for characterizing representations and computations." The extent to which the reader will find these theories convincing, however, depends largely upon her previous sympathies for specific eliminativist/reductivist claims. More generally, without the folk views held firmly at bay, it is hard to see these theories as genuine alternatives.[2] Good arguments for the eliminitivist claims—book 2 above—must first be given. Book 2, however, is not a part of *Neurophilosophy*.

II. THE QUESTION OF CONTENT

In this section, I want to talk about Churchland's theory of content. As many people have said, *Neurophilosophy* contains—at least, explicitly—very little about the subject. Why, then, discuss it? Most importantly, part III assumes the success of Churchland's philosophical view: the neurophysiological theories are largely suggestions about how to implement or realize that theory. In virtue of Churchland's view that the folk notion of propositional content will be replaced by a two-part theory of nonintentional representations and computational relations, the new scientific theory acquires its goal —namely, to supply "the foundations of a new paradigm for characterizing representations and computations." The theory of content sets the agenda.

On the Churchland theory of content (a view held and explicated by both Patricia and Paul Churchland[3]), folk psychology is viewed as having conflated two separate notions of content. On the one hand, there is "translational" content. Roughly speaking, this is "what the agent has in mind"—*how* the agent represents an external event to herself. So, for example, if I exclaim, "Zeus's thunderbolts!" (sincerely and with feeling), whenever the gas oven lights with a bang, the translational content of the thought is about Zeus. On the other hand, because that thought—in whatever guise—reliably correlates with the ignition of gas, its "calibrational" content is about gas explosions. That is, I think *of* the explosion (calibrational content) *as* an act of Zeus (translational content).

On the Churchlands' reductive view, the translational content of a neural state is identified with the "inferential/computational rela-

[2] For example, without first explaining in naturalistic terms (or explaining away) the intentionality of mental representations, it is hard to see how such states could possibly be "positions in phase space" (452). This issue is discussed in more detail below.

[3] See, for example, Paul M. Churchland, *Scientific Realism and the Plasticity of Mind* (New York: Cambridge, 1979); and P. M. and P. S. Churchland, "Functionalism, Qualia, and Intentionality," *Philosophical Topics*, XII (Spring 1981): 121–145.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tions it bears to all the rest of (the agent's) representations." Just as a linguistic term has meaning in virtue of its place within the "semantic net," so too is the meaning of a mental state a function of its inferential/computational relations to other representations. And representations—for one must now discharge the remaining intentional term—are those neural states with calibrational content. They are those states which reliably correlate with some property of the environment. Both theories of "meaning," of computation and of "representation," then, are placed squarely within the intellectual domain of neuroscience.[4]

Let me take a moment to put this theory into philosophical context. Despite the theory's origins, it is clear that the Churchland view differs significantly from those of Quine and Sellars. Quine, as we know, took the failure of the analytic/synthetic distinction plus the indeterminacy of translation to show that we ought to dispense with propositional attitudes and their semantic kin—or rather, that we would be wise to do so in our attempts to "limn the true and ultimate structure of reality." If there is no empirical solution to the problem of translation, then there is no sense in speaking of "a fact of the matter" about meaning—no phenomenon, "meaning," which stands in need of an explanation over and above that given by a theory of translation. Sellars' view is no less revisionary. If we are to regard our lore about mental states as fundamentally dependent on an analogy to linguistic practice, then it is open to question whether any "linguistic" properties are rightly attributable to the mental—and this includes both the notions of meaning and reference.

Contrast this with the Churchland view. It follows from the above that at least some neural states are meaningful, they yield a determinate semantic content, can be assessed for "synonymy," and are both representational and "referential."[5] Clearly, the Churchlands

[4] One important caveat should be noted. The Churchlands do not think that, thus, the folk psychological notion of content is reduced to neural/functional statements. Along with Stich, the Churchlands hold a "you have what I have" theory of content ascription. When I ascribe the belief that p to another agent, I assert (tacitly) that she possesses a "semantic analogue" to whatever state I have when I believe that p. I assert that she has a neural state that plays a cognitive/inferential role similar to some state in my own representational "economy." Nonetheless, a folk ascription does not serve to reliably "point at" some type of neural state. First, content ascriptions are both contextually sensitive and vague; second, folk ascriptions conflate calibrational with translational content (i.e., ascribe content on "nonindividualistic" grounds). The end result is a set of contradictory ascriptions —an inconsistency within folk theory that the Churchlands believe will preclude intertheoretic reduction.

[5] Although the conclusions of the Churchland theory appear thoroughly un-Quinean, the roots of the divergence are to be found in a disagreement about the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

believe that we need a theory of meaning—that something needs to be explained—and they want to do so using largely traditional semantic concepts. They are not attempting, however, to explain the folk notion of content, to say how, in the naturalistic order of things, certain neural states come to have a particular propositional content. Theirs is a theory of "meaning"—but not of propositional content ascription.

What, then, are the phenomena that the Churchlands want explained and, assuming the robustness of the phenomena, how good are those explanations?

Let me begin with the notion of calibrational content. Here the phenomenon is not intentional states. Rather, what calibrational states—"representations"—explain is simply how we get about. Churchland writes,

> I take it as obvious that if there were no systematic relations between the external world and states of the brain, the animals could not survive or if they did it would be a miracle in the strict sense of the word. An owl would not know where the mouse is and he would not intercept it as it runs.[6]

There is not much to quarrel with in this quotation. I want to point out, however, that there is a significant difference between a systematic relation and the kind of correlational (representational) state to which the Churchlands appeal in their explanation of content. Imagine, for example, a simple creature with a digital instrument that registers set increases in light intensity. The function of the instrument is to control the amount of light into the eye, which can be done either with "shades" (i.e., filters) or by closing the lens aperture of the "eye." It works as follows. Each registered increment in intensity causes a filter to be lowered over the "eye." When all the ten filters have been used successively (i.e., ten intensity increases have been registered), the "eye" aperture clicks down one notch. The filters go back to starting position, the intensity meter falls back to zero, and the process begins again, with the ten filters being lowered over a new aperture. For this creature, the light intensities in the world cause systematic state changes in the light sensor, but the increments on the "meter" do not correlate with any particular light

---

proper method of translation. This way of framing the disagreement was suggested by Patricia Churchland in Paul Churchland's *Scientific Realism and the Plasticity of Mind*, sect. 9, pp. 63–74.

[6] "Replies to Comments" from "Symposium on Patricia Smith Churchland's *Neurophilosophy*," *Inquiry*, XXIX (June 1986): 139–273, p. 260.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

intensity. Once the meter is reset to zero, its settings (alone) no longer indicate the actual value of light intensity. Functionally speaking, then, this value does not matter to the "motor" mechanism. What this organism has, in other words, is a simple "servo-mechanism."

Systems of the above kind have systematic relations with the environment. But they do not function in virtue of correlations qua correlations. Contrast this with, say, our best guess about one function of the retinotopic map. Given that light travels in straight lines, any given place on the retina correlates with a spot in visual space. A place on the retina "represents" a place in the world. If we have guessed right, these co-ordinates are used, qua co-ordinates, for various tasks, say, to guide motor movements. For example, on one version of the tensor-network theory, movement is made possible by functions that map coordinates encoded in sensory space onto co-ordinates in motor space. In such systems, it is states qua representations (unchanging correlations) and transformations between co-ordinates that do the explanatory work.

Why does this matter? Most anyone would agree with the conjecture that the brain has systematic relations with the environment. But that these relations consist in functional correlations is a much stronger empirical hypothesis. The question is: Which view does Churchland endorse? If the former weaker view, then our understanding of "content" is not furthered, for "content" requires an indicator relation. Nor does postulating systematic relations give neuroscientists any mandate for future research—it does not lay down a theoretical basis for explaining how we get about.

If Churchland endorses the stronger thesis,[7] however, there is a clear link with the question of content—functional correlations with external properties provide the neural analogue of reference. Still —and this is what I want to emphasize—the representational view is a strong empirical hypothesis with ready alternatives. If, as a matter of empirical fact, the brain manifests "servo" type mechanisms *and* "representational" type mechanisms *and* mechanisms for keeping track of and comparing internal states *and* mechanisms of functional kinds we have yet to discover, then very few of the brain's states and processes can be characterized by Churchland's theory. My question here is not whether to call these various functional types "representations." I do not think it much matters. Rather, I am arguing that, if the basis of brain function is not of the kind the Churchlands sug-

[7] This would seem to be the case. All the vector-tensor systems suggested by Paul Churchland (for such diverse sensory tasks as color vision, olfaction, face recognition, and vision) are of this form.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

gest, then we would not be able to rest our explanation of content solely upon this one kind of function, the "representational" state.

What, then, of "translational" content? Here the phenomenon to be explained is the "theory-laden" nature of mental states. As Paul Churchland has put it,[8] this is the problem of how the semantic content of a state can diverge from its causal couplings to the world —how the brain interprets its signals as representing one property rather than another. What, for example, is going on such that I see the gas explosion as an act of Zeus? Why might someone else see that same explosion as an explosion? As I said above, the Churchlands' answer is in terms of computational relations: for a representation to be theory-laden is for it to occupy a particular place in the computational "web."

To judge this answer, take a simple neurological example, that of the vestibular-ocular reflex (see *Neurophilosophy,* pp. 433–4441 for a discussion and competing theory thereof). The problem that the VOR solves is how to keep the eyes fixated on one part of the world despite head movements—it must rotate the eyes in the opposite direction of the head movement but at the same angular velocity. In understanding this process, David Robinson[9] has postulated a *neural integrator* ("integration" in the mathematical sense). First, the signal, which emerges from the vestibular system (semicircular canals), encodes the velocity of the head movement in the plane of the semicircular canals. This signal is relayed directly to the motor neurons. But in order to move the eyes, both the desired velocity and the desired deviation—position—are needed. Hence, a second parallel line goes to the *prepositus* nucleus where, through integration, position is calculated from velocity. This information is then relayed on to the motor neuron for a completed output signal to the eye muscles.

Given that neurological computations of this sort will probably prove quite common in brain function, how does understanding this kind of neural system clarify the nature of "theory-laden" states? At first the signal encodes velocity information; after integration, it encodes position. But what is velocity represented *as* at any given

[8] "Semantic Content: In Defense of a Network Approach," a reply to Kenneth Sayre's "Intentionality and Information Processing: An Alternative Model for Cognitive Science," *Behavioral and Brain Sciences,* IX, 1 (March 1986): 121–155, p. 141.

[9] For a review of this literature see his "Control of Eye Movements," in *Handbook of Physiology, The Nervous System,* II, 2 (Baltimore: Williams & Wilkins, 1981): 1275–1320; for an informal lecture by the same author, see "The Windfalls of Technology in the Oculomotor System," in *Investigative Opthamology and Visual Science,* XXVIII (December 1987): 1905–1924.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

time? Is it represented as velocity ("change in distance as a function of change in time"?) or as something else, and if so as what? Even for this simple system, it is unclear which are the telling facts.

There is, of course, something unfair about this way of posing the question. If the Churchland theory does not attempt to ascribe propositional content to neural states, we cannot in fairness demand to know what velocity, or anything else, is represented as. We cannot demand to know the propositional content of the individual neural states. The very goals of the theory preclude this line of criticism.

Nonetheless, there is, by the Churchlands' lights, a phenomenon to be explained. What that phenomenon is, however, is none too clear. On the initial phrasing of the problem, we are trying to understand the "theory-ladenness" of neural states, how the content of a neural state "diverges" from its causal origins. But there is, of course, no physical thing, the meaning, which is carried about, divided, and reshaped by the internal workings of the brain. In more general terms, we might ask how a neural state "comes to have meaning for its owner" or how a stimulus is "interpreted" by the brain. Again, these phrases are largely metaphorical. We do not have any clear sense of what neural "interpretation" or "meaning" amounts to. The task for the reductivist/eliminativist, then, is to understand what processes, if any, are pointed at by these (quasi) folk phrases.

Referring to computational relations as a whole does not solve the problem. Assuming that there is a principled physiological difference between you and me, when I think, "Zeus's Thunderbolts!", and you think, "Gas Explosion!", the explanation of our capacities will advert to (among other things) various principles of functional organization. Whereas the difference at hand may be explained by the particulars, the general ability to "represent as" will be understood with reference to functional kinds, common to different systems, at various levels of organization. The reason, then, that the VOR example appears so empty is that we have only the particulars. Given no principles, we are no closer to unpacking the folk psychological notions.

What the above problems with the theory of content illustrate is that explaining and making plausible an eliminativist/reductivist view is an enormous task. Although *Neurophilosophy* does not supply the details, it contributes a vantage point from which to define that project. Being the first and leading step, we should acknowledge it as such.

University of Illinois/Urbana-Champaign KATHLEEN A. AKINS

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

How Experiments End. PETER GALISON. Chicago: University of Chicago Press, 1987. xii + 330 p. Cloth $39.95, paper $15.95.

Peter Galison is in the philosophy and physics departments at Stanford University, and in its program for the history of science. His book tells how high energy physics changed from desktop devices into "big science." His three deeply researched central chapters of case histories (pp. 21–241) will form one of the classical accounts of that epoch, but the last two chapters of the book are a contribution to epistemology. I shall briefly mention a few of their lessons, warning that I shall be giving one philosopher's heightening of Galison's more cautious exposition.

1. *The title* is significant: How *do* experiments end? That includes the question of what makes a body of data compelling grounds for settling an issue? The question has long been central to empiricist attitudes to knowledge, but it has seldom been asked so cleanly. Of course, most experiments that do not end in disarray end by being salvaged: the workers get enough to continue their careers. But what about those which end by finding out something people want to know, and which become benchmarks for future enquiry? It is these, so we are taught by most philosophies, which are among the foundations of scientific knowledge. Experiments end in consequence of interactions among a small number of people, the instruments that they construct, and something to which they apply their apparatus. Hans Reichenbach called that the context of discovery; the results of an experiment were, however, evidence that provides a logical and ahistorical justification for some conclusions. A decade ago, "social construction of scientific facts" philosophers implied that there is no such distinction between justification and discovery, and that evidence is a social product; experiments end when people have worked out their differences. Galison is neither Reichenbachian nor constructionalist. He denies that evidence has a purely logical content used in justification. Not only are data produced in material circumstances, but what counts as evidence is the product of historical traditions of experimentation and instrumentation. But there are strong nonsocial determinants of inquiry. Galison aims at characterizing these constraints.

2. *Instrumental traditions:* "Instruments have a life of their own." Experimental and instrumental traditions may be largely independent of higher level theory—a big change in our beliefs about the world may leave many of the instruments intact, and there can be radical innovation in apparatus not connected with a scientific revo-

0022-362X/90/8702/103–106 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

lution. What counts as answering a question, and hence what serves to bring experiments to an end, will differ between traditions. In nuclear physics, one tradition counted events in the aggregate (and hence built machines that made myriad events and counted them), while another produced visual displays of single events. The one sought a compelling array of data en masse (and what makes it compelling is part of another tradition, namely, that of statistical analysis). The other wanted the visual presentation of a "golden event" convincingly (and literally) demonstrating what is true. These are fundamentally different conceptions of evidence. The elaboration of one kind of apparatus produces new conceptions of what counts as evidence; the idea of "evidence" and the physical means to produce it are constantly modified. "Evidence" ceases to be a purely logical concept, arbiter of what it is reasonable to believe. Data have status only within a set of practices engineered to produce them.

3. *Experimental structure:* Galison cites Clerk Maxwell on scientific apparatus (1876). The greatest theoretician/experimenter of the nineteenth century was no slouch at philosophical analysis. He showed how the experimenter isolates the apparatus from "disturbing agents" and creates phenomena that do not occur in nature until people have separated them from the prehuman course of events. He thought one did this by controlled transfers of energy, and he had a tidy classification, with three main heads, of the functions of an instrument: (1) a source of energy, (2) a transporter of energy, and (3) a measurer of the effects of the energy. This captures presciently much of the most celebrated experimental physics of the twentieth century, with one exception. The experimenter once related to theory through the apparatus and the results recorded in (3), processed in a minimal way. Then you could see what it all "meant." Increasingly we now add (4), procedures for analyzing, processing, and communicating data, procedures radically transformed by fast computation, not just at the level of sums, but also, for the visual tradition, by scanning and image enhancement. It is a commonplace to say that we live in the computer age, but no one has so succinctly argued that this modifies the structure of an experiment, not just by adding a new layer of workers, but by changing the very objects that count as the "data" for theoretical analysis.

4. *Scale:* Does it matter to epistemolology? Page-long lists of authors of a single paper grab attention. High cost determines bureaucratic review procedures for funding. There is a division of labor, with designers of detectors in a high-energy project being a different class of people—and using different physics—from designers of energy sources. There are echelons of industrial managers and safety

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

engineers, as well as the aforementioned data processors. Does this have any effect on scientific method and inference? Galison suggests that particular traditions of experimentation become much more rigid. With a hand-made desk-top device it was easy to modify any part of the instrument, to change it in significant ways when one revised one's idea of how the thing works. But the vast investment of big science in industrial plants—with a very small number of employees, despite the growth in number of Ph.D. physicists and their support staff—preselects admissible interfaces with phenomena. The class of possible data is settled by styles of instrumentation and by existing apparatus to a greater extent than science ever knew before. Big science builds into itself certain conceptions of the world and is almost "revolution-proof."

5. *Constraints:* What brings an experiment to an end? Galison distinguishes three levels of constraint on experiments: long-, middle-, and short-term. Some long-term constraints are theoretical, including metaphysical views of what the world could possibly be like. Galison urges that there are also long-term constraints arising from instrumental traditions, each demanding that only a certain kind of result is pertinent. Thus, the physicists who start with counters and scintillators develop the spark chamber and the wire chamber, new apparatus that will provide data different in kind from cloud chambers, bubble chambers, and nuclear emulsions. Thus, we step "down from the aristocratic view of physics that treats the discipline as if all interesting questions are structured by high theory."

His middle- and short-term constraints include some familiar ones, but also new insights. At the middle level, he emphasizes how theoretical assumptions can be *built into the apparatus itself.* Machines are not neutral probers of nature—and Galison has detailed examples to prove this. At the short-term level, he singles out, for example, the principles for routinely selecting which data to analyze, and reminds us that, at CERN, groups of physicists stand around the table "stamping IN and OUT on event candidates." This is neither rigid routine nor anarchy, but rather a craft skill, taught by apprenticeship within an experimental culture.

6. *Realism:* Galison might be read as fuelling the flames of "social construction" theories of knowledge. I think not. Usual accounts of scientific method, be they rationalist or empiricist, held that theories are true or false and that we judge of their truth value by their ability to predict and explain phenomena. Skeptical theories of a conservative and old-fashioned bent—instrumentalism, constructive empiricism—slightly changed the terms in this equation, talking of theories being adequate rather than true, that is, adequate to the phenomena,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

whose objectivity and identity was never put in question. The new skepticism—social constructionalism—challenges the presupposition that phenomena investigated experimentally exist independently of social interchange. Hence, it holds that theories never have any "retroactive" truth; we do not discover the truth so much as invent and maintain it. Galison, as I read him, also treats data and even phenomena as material productions of certain experimental and instrumental traditions. It is not just that we would obtain different data in different traditions, but that what counts as data is determined by the traditions of which he has given examples. This is not, however, a form of skepticism, because Galison is keenly aware of the realities about getting anything at all out of apparatus. The real meshing of experiment with the world occurs not by finding out which theories are true to the phenomena, but by finding out the ways in which the instrumental traditions in which they are made incarnate are able systematically to interact with their materiel, subject to the three levels of constraint. We might call this technological realism, not meaning another high-falutin "scientific realism" but "realism" in the ordinary sense of the word, the realism of the person who knows that hardly anything can be made that works, and that our visions of the world's microstructure are built around that humbling fact.

IAN HACKING

University of Toronto

Subject, Thought, and Context. Philip Pettit and John McDowell, eds. New York: Oxford University Press, 1986. 300 p. Cloth $49.95, paper $17.95.*

This book contains an introduction by the editors and nine new essays on issues in philosophy of language and philosophy of mind. The topics include the nature of psychological explanation, folk psychology, functionalism, the individuation of linguistic and intentional content, and Wittgenstein's views about private language. On the whole, the contributions are lucid, ambitious, and carefully argued. The book is recommended with enthusiasm.

* I have been helped considerably while writing this by several conversations with Anthony L. Brueckner.

0022-362X/90/8702/106–112

© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

I shall begin by discussing two of the papers at some length. Then, bowing to space limitations, I shall comment briefly on the others.

In a long essay entitled "Broad-minded Explanation and Psychology," Philip Pettit discusses two conceptions of the practice of explaining actions in terms of propositional attitudes. On the one hand, it is sometimes maintained that we explain actions by *regularizing* them—that is, by claiming that, in view of certain of the attitudes of their respective agents, actions can be seen as instances of nonaccidental regularities. On the other hand, attitude-based explanations are sometimes seen as attempts to *normalize* actions, where normalizing an action consists in showing that "in the light of the agent's intentional states it was required by a norm, in particular by a principle of behavioral rationality" (47). To borrow a phrase from John McDowell, we normalize actions by representing them as things that "rationally ought" to have happened, given the propositional attitudes of their respective agents.

In the opening sections of the paper, Pettit argues that it is in principle possible to give adequate regularizing explanations of actions, at the level of psychology, without employing the conceptual apparatus of propositional attitudes. Specifically, he maintains that it is possible to regularize actions by appealing to *narrow* psychological states (i.e., to states that can be individuated without making reference to extracranial objects and properties). It follows, he says, that, if we embrace the first of the foregoing conceptions of attitude-based explanations, then we will be forced to see attitude-based explanations as dispensable. Thus, according to Pettit, the first conception threatens to deprive the attitudes of their cognitive value.

In later sections of the paper, Pettit argues for the correctness of the second of the two conceptions, and also tries to show that it is much more hospitable to the attitudes than the first conception. Among other things, he maintains that the second conception gives us a guarantee that attitude-based explanations are *not* dispensable.

Pettit's paper is insightful and provocative. Like most authors who distinguish between regularization and normalization, however, he underestimates the position of those who hold that the task of normalizing an action is at bottom the same as the task of regularizing it in terms of a particular framework (viz., the framework of the attitudes). Moreover, it is not altogether clear that he is right in thinking that the arrival of a narrow psychology would detract from the cognitive value of the attitudes. Thus, even if it is possible to conduct the business of explaining behavior using only narrow coin, it could still be true that the attitudes are *there*. And it may well be that there are strong positive reasons for thinking that they *are* there. (It may even

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

be possible to establish their existence "constructively," by formulating an adequate causal or teleological theory of the nature of wide content.) Let us assume that it is possible to come up with such reasons, and let us assume also that there are good reasons for thinking that generalizations about the attitudes are more or less lawlike. (Recall that generalizations about the attitudes can support counterfactuals.) Under these assumptions, it seems that we should allow that attitudes can play a role in bona fide explanations of behavior—even if it is also possible to explain behavior in terms of narrow psychological states. By the same token, it seems that we should grant that the attitudes have a full share of cognitive value.

In a paper entitled "Cartesian Error and Perception," Tyler Burge undertakes to develop and defend his anti-individualist position concerning the individuation of intentional states. Burge opens by discussing several objections to his position which derive from Cartesian epistemological doctrines. The most interesting of these can be expressed as follows. According to anti-individualism, the individuation of one's intentional states depends upon relations one bears to one's physical and social environments. Thus, for example, if anti-individualism is correct, it is normally impossible to think that water is wet unless one has a mental representation that is causally connected with water in an appropriate way. It seems to follow that one cannot *know* that one is thinking that water is wet unless one has determined, by an investigation of the environment, that the appropriate causal relations obtain. As Descartes pointed out, however, our knowledge of our own intentional states is direct and authoritative. It seems to follow that anti-individualism is wrong.

As Burge points out, this objection presupposes that it is impossible to know that one is thinking that p unless one knows independently that the enabling conditions of thinking that p are satisfied (that is, unless one is independently aware of the conditions that must obtain in order for a thought to have the content that p, and one is also independently aware that these conditions actually do obtain). Burge rejects this presupposition. He does not give his reasons for rejecting it in the paper under review. In a more recent paper,[1] however, he offers a positive account of authoritative self-knowledge which implies that the presupposition is wrong.

In the second half of the paper, Burge tries to establish the main doctrine of anti-individualism, restricted to perceptual content. Specifically, he tries to establish the following claim: it is possible to fix

[1] "Individualism and Self-knowledge," this JOURNAL, LXXXV, 11 (November 1988): 649–663.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the intrinsic, nonintentional properties of a subject without thereby fixing the content of her perceptual representations. In the reviewer's judgment, Burge's argument for this claim is importantly different from his earlier arguments for related doctrines.

In discussing externalist accounts of perceptual content, Burge says a number of things which might be thought to provide the seeds of an argument against certain forms of skepticism. For example, he observes that, if an externalist account is true, then it "makes no sense to attribute systematic perceptual error to a being whose perceptual representations can be explained as the results of regular interaction with a physical environment" (131). It is tempting to attempt to turn this claim against the skeptic by arguing as follows: "It makes no sense to attribute systematic perceptual error to me. Hence, most of my perceptual beliefs are true. But my perceptual beliefs entail that I am not a brain in a vat. Hence, I am not a brain in a vat." (In stating this argument, I have assumed that it is appropriate to explain our perceptual representations as the results of regular interaction with an environment of some sort. This is conceded by most skeptical hypothesis, including the brain in a vat hypothesis and Descartes's evil genius hypothesis.)

To the reader's surprise, Burge provides no encouragement for the view that Putnamian arguments of this sort are sound. Why not? It is very much to be hoped that Burge will address this question in future work. [There is some evidence that Burge is disposed to endorse Brueckner's objection to Putnamian arguments.[2] According to Brueckner,[3] while Putnamian arguments can be used to establish the metalinguistic conclusion that the sentence 'I am not a brain in a vat' is true, they cannot be used to show that I am not a brain in a vat. This objection is extremely interesting, and it may well be that it points in the right direction. Advocates of the objection need to say more, however, about the relevant instance of the disquotation principle—that is, the instance which says that 'I am not a brain in a vat' is true if and only if I am not a brain in a vat. Contrary to what Brueckner maintains, it seems that I can claim to know that this instance holds in my idiolect without begging the question against the skeptic; for it seems that the instance holds, and that I can know that it holds, whether my language is English or vat-English (i.e., English with the interpretation it would have if used by a brain in a vat). In general, in order to know that '*S*' is true if and only if *S*, I need only know the meaning of the word 'true' and have reason to

[2] See *ibid.*, fn 6, p. 655.
[3] "Brains in a Vat," this JOURNAL, LXXXIII, 3 (March 1986): 148–167.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

believe that the embedded sentence has some determinate meaning or other in my idiolect. I need not have any independent information about the external determinants of the meaning of the embedded sentence. (Moreover, it would be a mistake to view my knowledge of the relevant instance of the disquotation principle as the trivial, metalinguistic knowledge that " 'I am not a brain in a vat' is true if and only if I am not a brain in a vat" *is true.* Rather, I can be said to know that 'I am not a brain in a vat' is true if and only if I am not a brain in a vat.[4])]

Gregory McCulloch's contribution, "Scientism, Mind, and Meaning," is concerned with many of the same issues as Pettit's paper. Thus, like Pettit, McCulloch wants to identify the sources of the cognitive value of folk psychology, and to show that its value would not be substantially diminished by the arrival of a psychology that trafficked only in narrow states. But there are some important differences. As we saw, Pettit argues for the importance and irreplaceability of folk psychology by maintaining that the explanations provided by a narrow psychology would have a different form, and also a different purpose, than explanations that appeal to intentional states. McCulloch takes a different line, arguing that the importance of folk psychology derives largely from its use in connection with enterprises that have little or nothing to do with explaining behavior. McCulloch also sketches an interesting position on the ontological status of propositional attitudes. This position likens the attitudes to secondary qualities, claiming that they are objectively real but nonetheless perspective-dependent ["in the sense that they are only had by objects in so far as they are thought of from certain points of view" (89)].

[4] In my judgment, Brueckner muddies the water by comparing his knowledge that 'Brueckner is not a brain in a vat' is true if and only if Brueckner is not a brain in a vat with his knowledge that 'Omega is not a regular cardinal' is true if and only if omega is not a regular cardinal. Since he does not know the meaning of 'Omega is not a regular cardinal', he says, it is necessary to view the second piece of knowledge as metalinguistic knowledge to the effect that " 'Omega is not a regular cardinal' is true if and only if omega is not a regular cardinal" is true. By the same token, he says, unless he has investigated the external determinants of the meaning of 'Brueckner is not a brain in a vat', it is necessary to view the first piece of knowledge as metalinguistic. There is an important difference here that destroys the analogy, however. Brueckner wants to say that he does not know the meaning of either 'Omega is not a regular cardinal' or 'Brueckner is not a brain in a vat'. But the sense in which this is true of the former sentence is quite different from the sense in which it is true of the latter. He does not know the meaning of the former in the sense that he does "not understand the technical terminology in the sentence" (165)—that is, in the sense that the sentence lacks a fully determinate meaning in his idiolect. This cannot be said of the latter sentence.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

In "Physicalist Thinking and Behavior," Jennifer Hornsby contrasts two notions of behavior, one that implies that behavioral phenomena are individuated by intentional and other relational properties, and another that represents behavior as consisting of colorless bodily movements. She maintains that the allure of functionalism is due in part to a failure to distinguish sharply between these two notions. Furthermore, she tries to show that, if philosophical accounts of mental states are to be materially adequate, then they must employ the first of the notions rather than the second.

McDowell's contribution, which is entitled "Singular Thought and The Extent of Inner Space," investigates externalist theories of intentional content, Cartesian epistemological doctrines, the Cartesian philosophy of mind, and some of the relationships between these topics. McDowell argues—intriguingly, if somewhat opaquely—that externalism can help to liberate us from some of the tentacles of Cartesianism. (There are interesting points of contact between this paper and Burge's essay.)

McDowell's essay is followed by a short piece by David Wiggins. This piece, which carries the title "On Singling out an Object Determinately," defends the determinacy of identity (i.e., the doctrine that it is always fully determinate whether $x = y$). Unfortunately, the paper is already somewhat dated. In view of recent essays by Terrence Parsons[5] and Peter van Inwagen,[6] it now seems that the case against the determinancy of identity is much stronger than it appeared to be in 1986, when the volume under review was published.

Next we find an essay entitled "What Determines Truth Conditions?" by Christopher Peacocke. This essay is reprinted as chapters 2 and 3 of Peacocke's *Thoughts*.[7] The reader is referred to reviews of Peacocke's book for discussion.[8]

Crispin Wright's "A Cogent Argument against Private Language?" proposes a new interpretation of Wittgenstein's private-language argument. Strikingly original and meticulously crafted, the paper can be hailed as a major contribution to Wittgenstein studies. There are philosophers, like myself, however, who deny that the traditional realist view of sensations leads to a doctrine of privacy, in anything like Wittgenstein's sense, and who therefore think that the

[5] "Entities Without Identity," *Philosophical Perspectives*, I (1987): 1–19.
[6] "How to Reason about Vague Objects," *Philosophical Topics*, XVI, 2 (Spring 1988): 255–284.
[7] New York: Blackwell, 1986.
[8] See, for example, John Campbell, "Critical Notice of *Thoughts*," *Mind*, XCVII (January 1989): 135–143.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

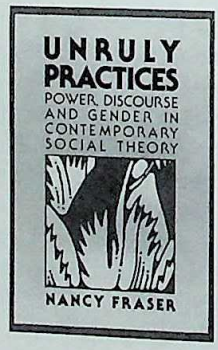
literature on privacy is, relative to substantive questions about sensations and the content of our actual sensory concepts, largely unmotivated. These philosophers will inevitably feel that, despite its brilliance, Wright's paper confirms their skepticism about the significance of Wittgenstein's argument.

Jonathan Lear's essay, "Transcendental Anthropology," explores two aspects of Wittgenstein's later conception of philosophy. One aspect, which Lear calls *anthropological,* comes to the fore in statements like 'What we are supplying are really remarks on the natural history of human beings'.[9] Such statements suggest that Wittgenstein means to be conducting an empirical or quasi-empirical inquiry. But there is also plenty of evidence that Wittgenstein means to be seeking a reflective, nonempirical insight into the conditions of applicability of concepts. This is the *transcendental* aspect of his thought. Lear is concerned with the tensions between these two aspects.

CHRISTOPHER S. HILL

University of Arkansas

[9] See Wittgenstein, *Philosophical Investigations,* G.E.M. Awscombe, trans. (New York: Blackwell, 1968), I, 415.



UNRULY PRACTICES

POWER, DISCOURSE AND GENDER IN CONTEMPORARY SOCIAL THEORY

NANCY FRASER

"A wonderfully rich and insightful collection of well-integrated essays on important current thinkers and social movements." *Martin Jay*

201 pages, $35.00 cloth, $14.95 paper

UNIVERSITY OF MINNESOTA PRESS
1-800-388-3863 VISA/MasterCard

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVII, NO. 3, MARCH 1990

DESCRIPTIVE PRONOUNS AND DONKEY ANAPHORA*

It is widely held that the following so-called "donkey sentences" create insurmountable semantical problems for traditional accounts of quantification and anaphora:

(1) Every man who bought a donkey vaccinated it.

(2) If John buys a donkey he vaccinates it.

(i) Gareth Evans[1] has argued that the anaphoric relation between 'a donkey' and 'it' in (1) cannot be captured by a theory that treats the subject expression 'every man who bought a donkey' as a "logical unit," i.e., as a restricted quantifier. (ii) Several philosophers and linguists have argued that Evans's own pioneering approach to pronouns and anaphora is undermined by the anaphoric relations in (1) and (2).[2] (iii) It has been argued that the same anaphoric relations thwart a unitary Russellian (i.e., existential) analysis of indefinite

* I am indebted to Irene Heim, James Higginbotham, Jeffrey King, Peter Ludlow, John Perry, and Scott Soames for extensive comments. Thanks also to Nuel Belnap, Jeremy Butterfield, Martin Davies, Gilbert Harman, Richard Holton, David Lewis, Trip McCrossin, Saul Kripke, David Magnus, Michaelis Michael, and Dan Sperber for helpful suggestions. The issues discussed here are treated more fully in my *Descriptions* (Cambridge: MIT, forthcoming), chs. 5–6.

[1] "Pronouns, Quantifiers, and Relative Clauses (I)," *Canadian Journal of Philosophy*, VII, 3 (1977): 467–536, and *The Varieties of Reference* (New York: Oxford, 1982).

[2] See, e.g., Irene Heim, *The Semantics of Definite and Indefinite Noun Phrases* (New York: Garland, 1988), pp. 44–102; Barry Richards, "On Interpreting Pronouns," *Linguistics and Philosophy*, VII, 3 (1984): 287–324; and George Wilson, "Pronouns and Pronominal Descriptions: A New Semantical Category," *Philosophical Studies*, XLV, 1 (1984): 1–30. Heim's book contains a lucid and comprehensive account of the recent history of donkey anaphora.

0022-362X/90/8703/113-150

© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

 Digitized by Arya Samaj Foundation Chennai and eGangotri

descriptions.[3] (iv) It has even been argued that such anaphoric relations warrant a "radical departure from existing frameworks" and "a major revision of semantic theory."[4]

By pulling together and refining some quite simple ideas that have appeared in the philosophy and linguistics literature, I want to suggest that sentences like (1) and (2) do not force us to abandon our standard conceptions of quantification and anaphora. In particular, I want to suggest that one can pursue the general approach to pronominal anaphora that Evans pioneered without abandoning either Bertrand Russell's semantics for indefinite descriptions or a general treatment of quantified noun phrases as restricted quantifiers. Evans's original theory is flawed in several ways; but some very natural modifications lead to a more interesting theory that provides plausible accounts of sentences (1) and (2) as well as some rather more complex examples.

I. WHAT IS THE PROBLEM OF DONKEY ANAPHORA?

On Russell's[5] account, indefinite descriptions (or "indefinites" for short) are devices of existential quantification. On this account, the logical form of (3) might be rendered as (3_1):

(3) Every man who bought a donkey was happy.

(3_1) $(\forall x)((\text{man } x \,\&\, (\exists y)(\text{donkey } y \,\&\, x \text{ bought } y)) \supset x \text{ was happy})$[6]

By the familiar equivalence between $\ulcorner(\exists x \phi) \supset \psi\urcorner$ and $\ulcorner(\forall x)(\phi \supset \psi)\urcorner$, where ψ contains no free occurrences of x, (3_1) is logically equivalent to:

[3] Hans Kamp, "A Theory of Truth and Semantical Representation," in J. Groenendijk, T. Janssen, and M. Stokhof, eds., *Truth, Interpretation and Information* (Dordrecht: Foris, 1984), pp. 1–41; Heim; and Wilson. The truth of this claim is also assumed by Jon Barwise, "Noun Phrases, Generalized Quantifiers, and Anaphora," in P. Gärdenfors, ed., *Generalized Quantifiers* (Dordrecht: Reidel, 1987), pp. 1–29.

[4] Kamp, p. 2; see also Barwise.

[5] *Introduction to Mathematical Philosophy* (London: Allen & Unwin, 1918), pp. 167–172. For detailed commentary, cf. Ludlow and Neale, "Indefinite Descriptions: In Defense of Russell" (forthcoming).

[6] The fiction that the unrestricted quantifiers $\exists$ and $\forall$ suffice for treating quantification in natural language will be dropped in section II. There appears to be another reading of (1) on which the existential quantifier introduced by 'a donkey' is given wide scope; but this is not the reading we are interested in here. (A wide-scope reading comes through more clearly in (i) Every man who saw a painting I bought was excited. It is important not to confuse the idea that the indefinite description in (i) may take wide scope with the idea that it admits of a semantically referential interpretation. As Saul Kripke has emphasized [in "Speaker's Reference and Semantic Reference," in P. A. French, T. E. Uehling, and H. K. Wettstein, eds., *Contemporary Perspectives in the Philosophy of Language* (Minneapolis: Minnesota UP, 1977), pp. 6–27], interpreting a definite or indefinite description as taking wide scope is not at all the same thing as interpreting it referentially.)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(3_2) $(\forall x)(\forall y)((\text{man } x \ \& \ \text{donkey } y \ \& \ x \text{ bought } y) \supset x \text{ was happy})$

Now consider (1):

(1) Every man who bought *a donkey* vaccinated *it*.

where 'it' is understood as anaphoric on 'a donkey'. Following P. T. Geach,[7] it is pretty generally agreed that, on its most salient reading, (1) is true if and only if every man who bought at least one donkey vaccinated every donkey he bought. And, again following Geach (and also W. V. Quine[8]), it is widely held that pronouns anaphoric on quantified noun phrases function as *bound variables*. We might therefore render (1) as (1_2):

(1_2) $(\forall x)(\forall y)((\text{man } x \ \& \ \text{donkey } y \ \& \ x \text{ bought } y) \supset x \text{ vaccinated } y)$

Now, despite the equivalence noted above, there is no first-order formula that is to (1_2) as (3_1) is to (3_2). The "closest" we get is (1_1):

(1_1) $(\forall x)((\text{man } x \ \& \ (\exists y)(\text{donkey } y \ \& \ x \text{ bought } y)) \supset x \text{ vaccinated } y)$

in which the final occurrence of y lies outside the scope of the existential quantifier that ought to be binding it. It would seem, then, that we cannot represent (1) with a formula that reflects an existential treatment of 'a donkey' (on the intended reading). Consequently, a uniform Russellian analysis of indefinite descriptions would seem to be thwarted.[9]

The existential treatment of indefinite descriptions works so well elsewhere that we should think very hard before giving it up or substantially modifying it. Not only does it work for indefinites that are subjects, or direct or indirect objects of simple sentences, it also works for indefinites in restrictive relative clauses, as in (3). But when there is anaphora on an indefinite contained in a relative clause, as in

[7] *Reference and Generality* (Ithaca: Cornell, 1962), pp. 126ff. Geach focuses primarily on conditionals of the form 'If any man buys a donkey he vaccinates it'. It should be clear that, contrary to what is claimed by, e.g., Norbert Hornstein [in *Logic as Grammar* (Cambridge: MIT, 1984)], the phenomenon under investigation is not something that arises only when the indefinite description requires a so-called "generic" or "quasi-generic" reading. This is the reason I have used sentence (1) rather than, say, 'Every man who buys a donkey vaccinates it', which apparently admits of a "quasi-generic" reading.

[8] *Word and Object* (Cambridge: MIT, 1960).

[9] Positing a full-blown ambiguity in the indefinite article will not help; there is simply no reading of, e.g., 'A man from London came to see me last night', or 'I met a doctor in Spain', in which the indefinite has universal force. The ambiguity theory is also susceptible to criticisms raised by Jaakko Hintikka and Lauri Carlson, "Conditionals, Generic Quantifiers, and Other Applications of Subgames," in F. Guenthner and S. J. Schmidt, eds., *Formal Semantics and Pragmatics for Natural Language* (Dordrecht: Reidel, 1978), pp. 1–36.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

 Digitized by Arya Samaj Foundation Chennai and eGangotri

(1), we must apparently resign ourselves to a treatment of the indefinite as a device of universal quantification.

We find a similar situation in conditionals. Consider:

(4) If John buys a donkey he is happy.

If we follow the common practice of taking a conditional to be composed of two sentences and a binary sentential connective, and then take the pronoun 'he' in the consequent of (4) to refer to John, we might assign to (4) the following logical form:

(4_1) $(\exists x)$(donkey x & John buys x) $\supset$ John is happy

(For present purposes, I propose to put aside the question of how well the material conditional fares.) By the familiar equivalence again, (4_1) is equivalent to (4_2):

(4_2) $(\forall x)$((donkey x & John buys x) $\supset$ John is happy)

Notice that the scope of the existential quantifier in the antecedent of (4_1) does not extend to the consequent; in a word, its scope is *clausal*. It should come as no surprise, then, that, if the consequent contains a pronoun that is anaphoric on the indefinite description in the antecedent, that pronoun will not be captured by the existential quantifier. Take:

(2) If John buys *a donkey* he vaccinates *it*.

If the indefinite is treated existentially and if 'it' is treated as a variable, (2) comes out as:

(2_1) $(\exists x)$(donkey x & John buys x) $\supset$ John vaccinates x

in which the final occurrence of x lies outside the scope of the existential quantifier that ought to be binding it. It is not difficult to find a sentence of first-order logic that gives us the right truth conditions:

(2_2) $(\forall x)$((donkey x & John buys x) $\supset$ John vaccinates x)

Again, the indefinite must apparently be treated as introducing wide-scope universal quantification.

But there is a prima facie absurdity in the claim that indefinite descriptions in certain subordinate structures—restrictive relative clauses and the antecedents of conditionals—have universal rather than existential import. As (3) and (4) make clear, it is just false that every indefinite description that occurs in such a position requires a universal interpretation. The problem seems to arise only when there is anaphora of the sort exemplified in (1) and (2).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

From the point of view of just capturing the truth conditions of (1) and (2) (and some related sentences), it will, perhaps, suffice to treat indefinite descriptions as introducing universal quantification when (and only when) they are constituents of certain subordinate structures. But the weakness of this approach is clear. Insofar as we are seriously engaged in the construction of a semantical theory for a learnable language, we must aim for more than such a limited brand of truth-conditional adequacy. We want a systematic deliverance of truth conditions, a theory that projects the truth conditions of sentences on the basis of the "meanings" of their parts and their syntactical structures. And, on this score, treating indefinite descriptions in (and only in) certain subordinate structures as devices of wide-scope universal quantification is, at best, a tottering first step. Such a treatment gives us no explanation of the apparent "universalization" in (1) and (2).

I want to explore the idea that the "universalization" of the indefinite descriptions in (1) and (2) is a logical illusion. In particular, I want to explore the idea that it is the anaphoric pronoun (rather than the indefinite description) that has a universal character. The existence of sentences like the following suggests that this might be a good way to proceed:

(5) Every man who bought *two or more donkeys* vaccinated *them*.

(6) If John buys *several donkeys* he vaccinates *them*.

Here again there is "universalization." Take (5); this is true just in case every man who bought two or more donkeys vaccinated every donkey he bought.[10] But we cannot capture this fact by treating 'two or more donkeys' as a wide-scope quantifier—universal or otherwise—that binds 'them'. If the quantifier were universal, (5) would be equivalent to (1), which it is not;[11] if the quantifier were 'two or more', (5) would mean that there are two or more donkeys such that every man who bought them vaccinated them, which it does not. All this suggests very strongly that, if we are to understand what is going on in (1), (2), (5), and (6), we need to think about the semantics of the anaphoric pronouns, not the semantics of their antecedents.

[10] The fact that we find "universalization" in sentences containing quantified phrases other than indefinite descriptions is noted by Gilbert Harman, "Deep Structure as Logical Form," in D. Davidson and G. Harman, eds., *Semantics of Natural Language* (Dordrecht: Reidel, 1972), pp. 25–47; see also James Higginbotham and Robert May, "Crossing, Markedness, Pragmatics," in A. Belletti, L. Brandi, and L. Rizzi, eds., *Theory of Markedness in Generative Grammar* (Pisa: Scuola Normale Superiore, 1981), pp. 423–444.

[11] This problem generalizes, of course, to all sentences of the form ⌜Every man who bought *n or more donkeys* vaccinated *them*⌝, for arbitrary *n*.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

II. RESTRICTED QUANTIFICATION

The inadequacies of the unrestricted unary quantifiers ∀ and ∃ for treating quantification in natural language are well-known. With the aid of the binary sentential connectives '⊃' and '&', restricted quantifications of the form ⌜All *F*s are *G*s⌝ and ⌜Some *F*s are *G*s⌝ can be expressed; but as Nicholas Rescher[12] has shown, there is no connective that will work for a unary quantifier designed to capture the semantics of 'most', for example. It is natural, therefore, to treat the determiners 'every', 'most', 'some', 'the', 'no', etc., as devices that combine, in some fashion, with two predicates to form a sentence. There are two standard ways of formalizing this idea, what I shall call BQ and RQ. According to BQ, formally a determiner is a *binary quantifier,* a device that combines directly with a pair of formulae to form another formula.[13] According to RQ, a determiner is a device that combines with a formula to form a restricted unary quantifier; and a restricted quantifier combines with a second formula to form another formula.[14] Consider the following sentence:

(7) Every man was happy.

Putting quantifiers inside square brackets and the formulae with which they combine inside parentheses, in RQ notation (7) comes out as (7_{RQ}), and in BQ notation it comes out as (7_{BQ}):

(7_{RQ}) [every x: man x](x was happy)

(7_{BQ}) [every x](man x:x was happy)

Notational variants, one would naturally think. Translation between formulae of the two systems looks easy enough, even for more complex examples such as (3):

[12] "Plurality Quantification," *Journal of Symbolic Logic,* XXVII, 3 (1962): 373–4. Furthermore, 'most *F*s' is not definable in first-order logic at all, even over finite domains [see, e.g., Barwise and Robin Cooper, "Generalized Quantifiers and Natural Language," *Linguistics and Philosophy,* IV, 2 (1981): 159–219].

[13] See, e.g., Evans, *op. cit.,* and "Pronouns, Quantifiers, and Relative Clauses (II)," *Canadian Journal of Philosophy,* 7 (1977): 777–797; David Wiggins, " 'Most' and 'All': Some Comments on a Familiar Programme, and on the Logical Form of Quantified Sentences," in M. Platts, ed., *Reference, Truth, and Reality* (New York: Routledge & Kegan Paul, 1979), pp. 318–346; and Martin Davies, *Meaning, Quantification, Necessity* (New York: Routledge & Kegan Paul, 1981). It should be remarked that these quantifiers are "binary" not in the standard sense that they simultaneously bind two distinct variables [see Andrej Mostowski, "On a Generalization of Quantifiers," *Fund. Math.,* XLIV (1957) 12–36], but in the sense that they combine with pairs of formulae to form formulae.

[14] See, e.g., Higginbotham and May, "Questions, Quantifiers, and Crossing," *The Linguistic Review,* I, 1 (1981): 41–80; and Barwise and Cooper.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(3) Every man who bought a donkey was happy.

(3_{RQ}) [every x: man x & [a y: donkey y](x bought y)](x was happy)

(3_{BQ}) [every x](man x & [a y](donkey y: x bought y):(x was happy))[15]

It would seem safe to assume, then, that once the syntax and semantics of BQ and RQ are fleshed out, the systems will be equivalent in expressive power and congruent in respect of semantical utility.

Evans[16] has contested this assumption. Again, it is the anaphoric relation in sentence (1) that apparently causes the trouble:

(1) Every man who bought *a donkey* vaccinated *it*.

According to Evans, (a) there are knock-down arguments for a sharp distinction between bound and unbound anaphoric pronouns, (b) the pronoun 'it' in (1) is unbound, and (c) if the pronoun 'it' in (1) is an unbound anaphor, the semantical connection between antecedent and anaphor can be forged in BQ but not in RQ. As we shall see, Evans is wrong on this last point. The only interesting difference between BQ and RQ is that the latter allows us to view quantified noun phrases like 'every man', 'every man who bought a donkey', etc., as syntactical and semantical units. Consequently, its formulae tend to be easier to parse, and for this reason RQ is the system I shall adopt.

For present purposes, only the barest outlines of RQ are necessary. Let us informally modify the formation rules of a standard first-order language by replacing the rules concerning $\exists$ and $\forall$ with the following:

(Q1) If ϕ is a well-formed formula that contains at least one occurrence of a term b and no occurrence of (the variable) x, and if D is a determiner, then $\ulcorner[Dx\colon \phi/x]\urcorner$ is a well-formed quantifier phrase (where $\ulcorner\phi/x\urcorner$ is the result of replacing at least one occurrence of b in ϕ by x).

(Q2) If ψ is a well-formed formula that contains at least one occurrence of a term b and no occurrence of x, and if $\ulcorner[Dx\colon \phi]\urcorner$ is a well-formed quantifier phrase, then $\ulcorner[Dx\colon \phi](\psi/x)\urcorner$ is a well-formed formula (where $\ulcorner\psi/x\urcorner$ is the result of replacing at least one occurrence of b in ψ by x).[17]

[15] The mechanisms involved in spelling out the semantics of restrictive relative clauses will be addressed in section VI.

[16] In "Pronouns, Quantifiers, and Relative Clauses (I)," pp. 136–9, and in *The Varieties of Reference*, pp. 58–9.

[17] In place of (Q1) and (Q2), BQ will have the following: (*) If ϕ and ψ are well-formed formulae, each of which contains at least one occurrence of b and no occurrence of x, and if D is a determiner, then $\ulcorner[Dx](\phi/x\colon\psi/x)\urcorner$ is a well-formed formula.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The notion of variable-binding operative in (Q1) and (Q2) is straightforward. We have two types of variable-binding operator, determiners and quantifiers. A determiner Dx binds every free occurrence of x in the formula with which it combines to form a quantifier. And a quantifier $\ulcorner[Dx\colon \phi]\urcorner$ binds every free occurrence of x in the formula with which it combines to form a formula. Using this notation, we can perspicuously and unambiguously represent the quantificational structures of sentences with more than one quantifier, the scope of such devices being understood exactly as in first-order logic. For example, the alleged ambiguity in (8) is captured by (8_1) and (8_2):

(8) Every boy danced with a girl.

(8_1) [every x: boy x]([a y: girl y](x danced with y))

(8_2) [a y: girl y]([every x: boy x](x danced with y))[18]

We can now write some sample truth clauses. If $|F|$ is the cardinality of F, and F is the set of things that are F, then:

(*1) $\ulcorner$[every x: Fx](Gx)$\urcorner$ is true iff $|F - G| = 0$

(*2) $\ulcorner$[no x: Fx](Gx)$\urcorner$ is true iff $|F \cap G| = 0$

(*3) $\ulcorner$[some x: Fx](Gx)$\urcorner$ is true iff $|F \cap G| \geq 1$

(*4) $\ulcorner$[an x: Fx](Gx)$\urcorner$ is true iff $|F \cap G| \geq 1$

(*5) $\ulcorner$[most x: Fx](Gx)$\urcorner$ is true iff $|F \cap G| > |F - G|$

[I have not attributed existential import to $\ulcorner$every $F\urcorner$; if it is wanted, add "and $|F| \geq 1$" to the right hand side of (*1).]

Following a suggestion made by Arthur Prior,[19] Russell's brilliant insight that definite descriptions are quantificational rather than

[18] By using the resources of a semantically well-understood and perspicuous notation, the semanticist is not committed to a notion of "Logical Form" (LF) construed as a level of syntactical representation. The work of Noam Chomsky, Higginbotham, May, and others suggests very strongly, however, that a variety of subtle facts about anaphora, quantification, and variable binding can be explained neatly on the assumption that there is such a level of syntactical representation. See Chomsky, *Lectures on Government and Binding* (Dordrecht: Foris, 1981); Higginbotham, "Pronouns and Bound Variables," *Linguistic Inquiry*, XI, 4 (1980): 679–708, and "The Logic of Perceptual Reports: An Extensional Alternative to Situation Semantics," this JOURNAL, LXXX, 2 (February 1983): 100–127; May, *Logical Form: Its Structure and Derivation* (Cambridge: MIT, 1985); and Higginbotham and May, *op. cit.* For an elementary exposition, see *Descriptions*, ch 5.

[19] "Is the Concept of Referential Opacity Really Necessary?" *Acta Philosophica Fennica*, XVI (1963): 189–198. The same suggestion is also made by Paul Grice, "Vacuous Names," in Davidson and Hintikka, eds., *Words and Objections* (Dordrecht: Reidel, 1969), pp. 118–145, and by Richard Sharvy, "Things," *The Monist*,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

referential is captured by treating 'the F' as a restricted quantifier $\ulcorner$[the x: Fx]$\urcorner$, i.e., by treating the determiner 'the' as a member of the same syntactical and semantical category as 'every', 'no', 'some', 'a', 'most', etc. According to Russell,[20] the logical form of $\ulcorner$the F is $G\urcorner$ is given by:

(9) $(\exists x)(Fx \,\&\, (\forall y)(Fy \supset y = x) \,\&\, Gx)$

As Russell points out, on this account, a definite descriptions is really a complex indefinite description, i.e., an existentially quantified noun phrase with a uniqueness condition built in. But we should not become preoccupied with the existential character of the analysis. As (9) makes clear, it is by introducing universal quantification that the implication of uniqueness is captured: $\ulcorner$the F is $G\urcorner$ is true if and only if (i) all Fs are Gs and (ii) there is exactly one F. In other words, definite descriptions are as universal as they are existential. And as Noam Chomsky[21] has pointed out, if we focus on universality, the relationship between singular and plural descriptions comes clearly into view. Intuitively, there is just a cardinality difference between the truth conditions of a sentence containing a singular description and those of the corresponding sentence containing the description in its plural form. For 'the F is G' to be true there must be exactly

---

LIII, 4 (1969): 488–504. Attempts to incorporate this idea into a general syntactical and semantical account of quantifiers in natural language can be found in Richard Montague, "English as a Formal Language," in Richmond Thomason, ed., *Formal Philosophy: Selected Papers of Richard Montague* (New Haven: Yale, 1974), pp. 108–221; and, in a more recent syntactical setting, in Higginbotham and May, "Questions, Quantifiers, and Crossing."

[20] "On Denoting," *Mind*, XIV, (1905): 479–493; and A. N. Whitehead and Bertrand Russell, *Principia Mathematica*, I (New York: Cambridge, 2nd ed., 1927), pp. 173ff. It is sometimes suggested that the theory of descriptions is too cumbersome and unwieldy to merit a place in a serious compositional semantics for natural language. Such a charge puts too much weight on the particular formalism of *Principia Mathematica*, and not enough on the genuine semantical insights of the theory itself. Russell's semantics for descriptions is quite compatible with the view that descriptions are restricted quantifiers (see below).

For the purposes of this paper, I am going to assume that Russell's theory of descriptions provides a more or less correct account of the semantics of definite descriptions in natural language on at least one of their uses. It is irrelevant to present concerns whether or not an additional semantically referential reading is required in order to make sense of cases involving so-called "referential" uses of descriptions. Following the lead of Grice and Kripke, in *Descriptions* I argue that referential usage is of no semantical import; nothing turns on this here. It should also be pointed out that endorsing Russell's theory of descriptions involves no commitment to a sense-datum epistemology or to a descriptive analysis of proper names.

[21] "Questions of Form and Interpretation," *Linguistic Analysis*, I, 1 (1975): 75–109.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

one F; for ⌜the Fs are Gs⌝ to be true there must be more than one F.[22] The following truth clauses make the relationship transparent:

(*6) Where 'F' is singular:

⌜[the x: Fx](Gx)⌝ is true iff $|F - G| = 0$ and $|F| = 1$

(*7) Where 'F' is plural:

⌜[the x: Fx](Gx)⌝ is true iff $|F - G| = 0$ and $|F| > 1$[23]

Let us now allow for the possibility of "numberless" or "neutral" descriptions, that is, descriptions that are, from a semantical perspective, neither singular nor plural, i.e., silent on whether $|F| = 1$ or $|F| > 1$. Although English nouns and verbs are typically marked for number, the language may well contain numberless descriptions. Phrases of the form 'whoever is F' are quite possibly such descriptions. As G. E. Moore[24] remarked, ⌜whoever is F is G⌝ entails nothing about uniqueness; it might be paraphrased as ⌜the person (or persons) who is (or are) F is (or are) G⌝, which makes clear its neutrality as to semantical number. Let us use ⌜[whe x: Fx]⌝ for a numberless description:

(*8) ⌜[whe x: Fx](Gx)⌝ is true iff $|F - G| = 0$ and $|F| \geq 1$

[22] It is well-known that some sentences containing plural noun phrases like 'Russell and Whitehead', 'a man and a woman', 'the men', 'three women', and so on admit of (or require) collective or group readings. The task of providing an adequate semantics for plural noun phrases that admit of such readings is clearly a very general task that has nothing to do with descriptions per se, and to that extent I shall not attempt to say anything significant about collective readings of descriptions here.

[23] To treat a singular definite description ⌜the F⌝ as a restricted quantifier ⌜[the x: Fx]⌝ is not to propose an alternative to Russell's theory; it is just to find a more congenial method of stating it. In effect, (*6) stipulates that ⌜[the x: Fx](Gx)⌝ is equivalent to ⌜$(\exists x)(Fx \,\&\, (\forall y)(Fy \supset y = x) \,\&\, Gx)$⌝. We are not, by virtue of being Russellians, committed to the view that the relevant sentences must be beaten into sentences of the language of *Principia Mathematica*. From the point of view of explicating logical structure, treating descriptions as restricted quantifiers, results not in a clash with Russell but in an explanation of where the theory of descriptions fits into a more general theory of natural language quantification, a theory in which determiners are treated as members of a unified syntactical and semantical category.

[24] "Russell's 'Theory of Descriptions'," in P. A. Schilpp, ed., *The Philosophy of Bertrand Russell* (New York: Tudor, 1944), pp. 177–225. Syntactically, of course, 'whoever' is singular, at least if verb agreement is anything to go by, witness the third person singular inflections on 'buys' and 'is' in, e.g., 'Whoever buys this is getting a bargain'. Apparently, some languages exploit the possibility of numberless descriptions on a far greater scale. Of course, if there are no singular or plural descriptions in these languages, it will not always be clear what counts as evidence for numberless descriptions, rather than systematic ambiguity between singular and plural descriptions. It is surely because English nouns (and verbs) are typically marked for number that we treat 'the sheep' as ambiguous rather than numberless.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The importance of numberless descriptions will become clear in sections VI and VII. Right now, we need to turn to pronouns anaphoric on quantifiers and to the relationship between pronouns and definite descriptions.

III. BOUND ANAPHORS

Following Geach and Quine, it is widely held that pronouns anaphoric on quantified noun phrases function like the bound variables of quantification theory. As Evans[25] has argued, however, not all pronouns with quantified antecedents can be treated in this way. Consider:

(10) John bought *some donkeys* and Harry vaccinated *them*.

If the pronoun 'them' is treated as a variable bound by 'some donkeys', the logical form of (10) will be:

(10_1) [some x: donkeys x](John bought x & Harry vaccinated x)

Evans points out that this is wrong on several related counts. First, (10_1) can be true even if Harry did not vaccinate all of the donkeys John bought, whereas (10) cannot [if John bought ten donkeys and Harry vaccinated only two of them, (10_1) would be true whereas (10) would not]. The bound proposal simply delivers the wrong truth conditions. As a corollary, it runs into a further problem that it will be fruitful to explore. A conjunction is true if, and only if, both of its conjuncts are true; so if (10) is true, so is (11):

(11) John bought some donkeys.

The bound analysis does in fact make the right prediction here, because (11_1), which represents (11), is a consequence of (10_1):

(11_1) [some x:donkeys x](John bought x)

But this is solely due to the choice of example. Consider the following:

(12) John bought *exactly two donkeys* and Harry vaccinated *them*.

(13) *Just one man* drank rum and *he* was ill afterward.

Take (13); if this sentence is true, then so is (14):

(14) Just one man drank rum.

But the bound analysis misfires here. If the scope of the quantifier

[25] "Pronouns, Quantifiers, and Relative Clauses (I)." See also his "Pronouns," *Linguistic Inquiry*, XI: 337–362.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

phrase 'just one man' is extended to the second conjunct, so as to capture the pronoun 'he', (13) will come out as (13_1):

(13_1) [just one x: man x](x drank rum & x was ill)

But (13_1) captures the truth conditions of (15):

(15) Just one man drank rum and was ill.

And unlike (13), the truth of (15)/(13_1) is quite compatible with the falsity of (14). [For example, (15)/(13_1) can be true if two (or more) men drank rum, but obviously (14), and hence (13), cannot.] Indeed, whenever a quantified antecedent is not monotone increasing the bound analysis will fail in this way.[26] With antecedents of the form ⌜no Fs⌝ the problem is especially acute. The bound analysis predicts that the incoherent (16) is equivalent to (17):

(16) John bought *no donkeys* and Harry vaccinated *them*.

(17) John bought no donkeys that Harry vaccinated.

Of course, if cross-sentential binding is prohibited, these problems do not arise. The upshot of all this is that, among those pronouns anaphoric on quantifiers, we need to distinguish between those which are bound (in the familiar sense) and those which are not.

There turns out to be a simple and precise syntactical constraint on when an anaphoric pronoun can be interpreted as a bound variable.[27] First, a very rough characterization: if P is a pronoun that is anaphoric on a quantifier Q, then P is bound by Q only if P is located inside the smallest clause containing Q. We can make this precise with the aid of some elementary configurational notions from contemporary grammatical theory. The first notion we need is that of dominance, which concerns the hierarchical organization of syntactical constituents. Consider the following (highly simplified) tree diagram for (18):

(18) Every man thinks (that) Mary loves him.

[26] Following Barwise and Cooper, a quantifier ⌜[Dx: Fx]⌝ is monotone increasing just in case the following is a valid inference:

[Dx: Fx](Gx)
[every x: Gx](Hx)

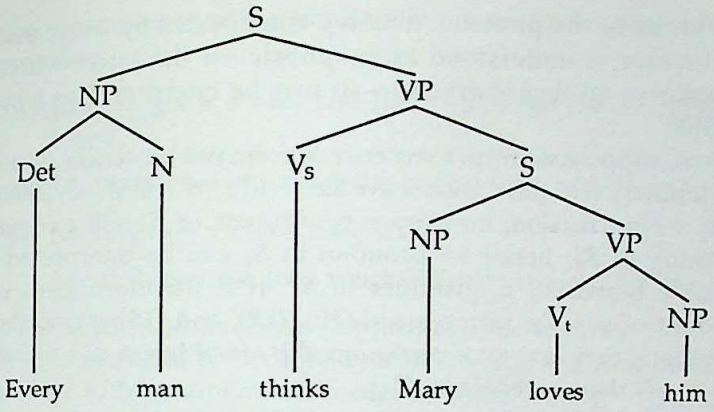
[Dx: Fx](Hx)

for arbitrary G and H.

[27] Indeed, as Higginbotham has remarked, ultimately a simple and precise syntactical characterization of Russell's notion of the scope of a quantifier. (To say that there is a syntactical characterization of scope is not, of course, to deny that scope is a semantical phenomenon.)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri



A node (or point) in a syntactical tree dominates its own constituents, that is, every node beneath it that can be traced back to it. Thus, in (18) the major verb phrase (VP) dominates the sentential verb (V_S) 'thinks', and everything in the embedded sentence (S) 'Mary loves him'. And the embedded verb phrase 'loves him' dominates the transitive verb (V_t) 'loves' and the noun phrase (NP) 'him'.

The central notion we need in order to characterize bound anaphora is that of *c-command:*

(P1) A noun phrase α c-commands a noun phrase β if and only if the first branching node dominating α also dominates β (and neither α nor β dominates the other).

By this definition, in (18) the subject NP 'every man' c-commands everything in the VP 'thinks Mary loves him'; in particular, it c-commands the pronoun 'him'.

We can now state the syntactical constraint on bound anaphora which Evans uncovers:

(P2) A pronoun P that is anaphoric on a quantifier Q is interpreted as a variable bound by Q only if Q c-commands P.[28]

[28] The 'only if' in (P2) can, I believe, be strengthened to an 'if and only if' (see *Descriptions*, ch. 5). To avoid distracting engagements, I have not used the stronger version here. For the purposes of this paper, I am making the simplifying assumption that (P2) is a constraint on surface syntax. As May and Higginbotham have stressed, however, in order to account for the anaphoric relations in examples like the following (due originally to Peter Geach and Benson Mates) (P2) must in fact hold at LF, a level of syntactical representation once removed from surface structure: (i) The man who bought *each donkey* vaccinated *it*. (ii) The father of *each girl* waved to *her*. (iii) The woman *every true Englishman* respects is *his* mother.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

In (18), since the pronoun 'him' is c-commanded by 'every man', if the former is understood as anaphoric on the latter—it could be understood demonstratively—it may be interpreted as a bound variable.

Now, suppose we form a sentence S_0 from two sentences S_1 and S_2, and a binary sentential connective like 'and', 'or', or 'if . . . then'. In such a construction, no proper constituent of S_1 will c-command anything in S_2; hence no pronoun in S_2 can be interpreted as a variable bound by a quantifier in S_1. (P2) therefore gives us an account of why the pronouns in (10), (12), and (13) are not bound anaphors: they are not c-commanded by (and hence not within the scope of) their antecedents.[29]

IV. DESCRIPTIVE ANAPHORS

We have established that the pronoun in (10) is not functioning as a bound variable:

(10) John bought *some donkeys* and Harry vaccinated *them*.

How, then, is the pronoun to be interpreted? A plausible paraphrase of (10) is (10_2):

(10_2) John bought some donkeys and Harry vaccinated *the donkeys John bought*.

This suggests to Evans that the unbound pronoun in (10) should be interpreted via the plural definite description 'the donkeys John owns', as what he calls an E-type pronoun. On this account, we might represent the logical form of (10) as (10_3):

(10_3) [some x: donkeys x](John bought x) &

[the x: donkeys x & John bought x](Harry vaccinated x)[30]

Now, what of singular pronouns in the same environment? Consider:

(19) John bought *a donkey* and Harry vaccinated *it*.

[29] Furthermore, they are not c-commanded by their antecedents at the syntactical level LF.

[30] Strictly speaking, Evans would not endorse (10_3) as giving the logical form of (10). On his final account, E-type pronouns do not go proxy for descriptions, but rather have their references fixed by them. [On the notion of fixing reference by description, see Kripke, *Naming and Necessity* (Cambridge: Harvard, 1980), pp. 54–5.] Evans explicitly entertains and rejects the proxy view, but in due course we shall see that the proxy view is superior to Evans's view. (The idea that unbound anaphoric pronouns might stand in lieu of definite descriptions goes back at least to Quine, pp. 102–3, 113.)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Again, a paraphrase of the pronoun in terms of a description seems appropriate:

(19_1) John bought a donkey and Harry vaccinated *the donkey John bought*.

We might, therefore, represent the logical form of (19) as follows:

(19_2) [an x: donkey x](John bought x) &

[the x: donkey x & John bought x](Harry vaccinated x)[31]

A few brief remarks about this proposal are in order.

(i) It is intended to be a general theory that treats all unbound pronouns with quantified antecedents in the same way. Thus, the pronouns in each of the following sentences will be cashed out as descriptions constructed from the clauses containing their antecedents:

(12) *Just one man* drank rum and *he* was ill afterward.

(20) John found *several minor mistakes* in his proof, but he managed to correct *them* without too much difficulty.

(21) *A few professors* came to the party. *They* seemed to enjoy themselves.[32]

(22) *The inventor of bifocals* was a genius; *he* ate a lot of fish.

(23) *The women who came to the party* were irritated by Bill; *they* complained, in particular, about his chauvinism.

In order to provide an explicit account of these facts as well as some more complex ones we shall consider shortly, we need some

[31] It is also open to argue, along with Kripke (in "Speaker's Reference and Semantic Reference," p. 27, fn. 32) and David Lewis [in "Scorekeeping in a Language Game," *Journal of Philosophical Logic*, VIII, 3 (1979), pp. 339–359], that some unbound pronouns anaphoric on definite and indefinite descriptions may refer to individuals raised to salience by the utterances of the sentences containing their "antecedents." (This idea was also suggested by Grice in unpublished work.) It is not necessary to see such a proposal as conflicting with the descriptive proposal. One interesting idea (due to Kripke) is that unbound pronouns anaphoric on descriptions that are used referentially may be interpreted either referentially or descriptively. (As Kripke stresses, to say that a definite or indefinite description is used referentially is not to say that the description is interpreted as semantically referential, i.e., interpreted as a referring expression.) Either way, it is not necessary to agree with P. F. Strawson [*Introduction to Logical Theory* (London, Methuen, 1952), pp. 185ff.] that definite and indefinite descriptions functioning as the antecedents of unbound anaphoric pronouns must be semantically referential.

[32] In this example, the reflexive pronoun 'themselves' is bound by the descriptive pronoun 'they'.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

terminology. Recall that definite descriptions, whether singular or plural, can be treated as (restricted) universal quantifiers with specified existential import: ⌜the F is G⌝ is true if and only if all Fs are Gs and there is exactly one F; ⌜the Fs are Gs⌝ is true if and only if all Fs are Gs and there is more than one F. As shorthand for this, let us say that the definite article is used to signal maximality, which in the case of a singular description is uniqueness. More generally, let us say that:

(P3) A quantifier ⌜$[Dx: Fx]$⌝ is maximal if and only if ⌜$[Dx: Fx](Gx)$⌝ entails ⌜[every x: Fx](Gx)⌝, for arbitrary G.

On this account, quantifiers of the form ⌜the F⌝, ⌜the Fs⌝, ⌜each F⌝, ⌜each of the Fs⌝, ⌜every F⌝, ⌜all Fs⌝, etc., are maximal. Second, let us say that:

(P4) The *antecedent clause* for a pronoun P that is anaphoric on a quantifier Q occurring in a sentence ϕ is the smallest well-formed subformula of ϕ that contains Q as a constituent.

We might now put forward the following generalization:

(P5) If x is a pronoun that is anaphoric on, but not c-commanded by a quantifier ⌜$[Dx: Fx]$⌝ that occurs in an antecedent clause ⌜$[Dx: Fx](Gx)$⌝, then x is interpreted as the most "impoverished" definite description directly recoverable from the antecedent clause that denotes everything that is both F and G.[33]

More usefully, (P5) can be spelled out as the conjunction of ($P5_a$) and ($P5_b$):

[33] (P5) has its origins in the generalizations offered by Evans ["Pronouns, Quantifiers, and Relative Clauses (I)," p. 111] and by Davies (p. 171), neither of which I find entirely satisfactory. The most important differences are the following: (i) Evans treats descriptive anaphors as having their references fixed by description rather than as going proxy for descriptions. For reasons given by Davies and by Scott Soames ["Review of Gareth Evans's *Collected Papers*," this JOURNAL, LXXXVI 3 (March 1989): 141–156], and recapitulated later in this section, this view must be rejected along with the "weaker" view that descriptive pronouns are equivalent to wide-scope descriptions; thus (P5) is stated in such a way that it does not preclude scope interactions between descriptive pronouns and other operators. (ii) Evans and Davies both treat determiners as binary quantifiers (Evans claims that quantified noun phrases cannot be represented as restricted quantifiers because of donkey anaphora; see section V). (iii) Evans treats unbound pronouns that are anaphoric on definite descriptions as "pronouns of laziness" rather than as E-type pronouns; (P5) is stated in such a way that it is applicable in cases where the relevant pronouns are anaphoric on singular and plural definite descriptions; if Davies's generalization is applied to such cases, it can yield incorrect results. In cases where the antecedent is not maximal, (P5) is, I believe, equivalent to Davies's generalization, *modulo* my use of restricted quantifiers. I am here indebted to Ludlow and Soames for valuable discussion.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(P5$_a$) If x is a pronoun that is anaphoric on, but not c-commanded by a nonmaximal quantifier $\ulcorner[Dx\colon Fx]\urcorner$ that occurs in an antecedent clause $\ulcorner[Dx\colon Fx](Gx)\urcorner$, then x is interpreted as $\ulcorner[\text{the } x\colon Fx \,\&\, Gx]\urcorner$.

(P5$_b$) If x is a pronoun that is anaphoric on, but not c-commanded by a maximal quantifier $\ulcorner[Dx\colon Fx]\urcorner$ that occurs in an antecedent clause $\ulcorner[Dx\colon Fx](Gx)\urcorner$, then x is interpreted as $\ulcorner[\text{the } x\colon Fx]\urcorner$.

(ii) Evans explicitly rejects the view that E-type pronouns go proxy for definite descriptions in favor of the view that they have their referents fixed by description. As Evans puts it, an E-type pronoun refers to "those objects which verify (or that object which verifies) the sentence containing the antecedent quantifier."[34] Evans's main reason for rejecting the proxy view in favor of the reference-fixing view is that he thinks E-type pronouns do not give rise to the sorts of scope ambiguities that overt descriptions delight in when they interact with other operators. According to Evans, if we take a sentence S containing an E-type pronoun P and substitute for P the favored description, the resulting sentence S' may exhibit a scope ambiguity that is not exhibited by S. Consider attitude contexts. On the proxy view, (24) will come out as (25):

(24) *A man* murdered Smith, but John does not believe that *he* murdered Smith.

(25) A man murdered Smith, but John does not believe that the man who murdered Smith murdered Smith.

Thus, it delivers two readings for the clause containing the anaphoric pronoun (the so-called *de re* and *de dicto* readings) according as the description delivered by (P5) is given wide or narrow scope:

(24_1) [the x: man x & x murdered Smith]

(John does not believe that (x murdered Smith))

(24_2) John does not believe that

([the x: man x & x murdered Smith](x murdered Smith)).

Evans points out that it is natural to interpret (24) as attributing to John "merely a noncontradictory belief of the murderer that he is not the murderer."[35] (This is the reading delivered by his own formal theory.) On the proxy view, this attribution is captured by reading the second conjunct as (24_1).

[34] "Pronouns, Quantifiers, and Relative Clauses (I)," p. 111.
[35] *Ibid.*, p. 133.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

It is (24_2) that Evans finds troubling. As he points out, (24_2) attributes to John the self-contradictory belief that the man who murdered Smith did not murder Smith. The fact that this unlikely reading emerges on the proxy theory suggests to Evans that the correct analysis of descriptive pronouns is that they are rigid designators that have their referents fixed by description, an analysis which prevents them from being interpreted as if they took narrow scope.

There are three points here. First, although Evans claims that E-type pronouns are rigid designators whose references are fixed by description, we must really construe him as making the ("weaker") claim that they are equivalent to wide-scope descriptions—which is, in fact, the way they are treated in his semantical formalism.[36] The reason for this is that being rigid and being equivalent to a description that insists upon wide scope do not amount to the same thing.[37] And as Soames points out, typically E-type pronouns are not rigid, because at different circumstances of evaluation the clauses containing their antecedents will be verified by different objects; in which case the pronoun will refer to different objects (or a different object) at different circumstances. Consider (10) again:

(10) John bought *some donkeys* and Harry vaccinated *them*.

If 'them' were rigid, it would refer to the donkeys John actually bought—let us say, Eeyore and Dinah—in which case it would be true at any circumstance of evaluation in which Harry vaccinated Eeyore and Dinah as long as John bought some donkeys there, though not necessarily Eeyore and Dinah.

Second, it is not clear that (24) raises a genuine problem for the proxy theorist at all. It is open to argue that the *de dicto* readings of the anaphor clauses in both (24) and (25) are technically available, but equally unlikely. Third, there are examples of the same general form as (24) in which the *de dicto* reading of the clause containing the anaphoric pronoun is clearly available, if not preferred.[38]

(26) *A man* murdered Smith. The police have reason to think *he* injured himself in the process.

(27) Hob thinks that *a witch* killed Trigger. He also suspects that *she* blighted Mathilda.

[36] This point and the next are made by Soames. At one place Evans does say that "E-type pronouns are like descriptions which insist upon widest scope" ["Pronouns, Quantifiers, and Relative Clauses (I)," p. 132, fn. 67].

[37] For discussion, see the 1980 preface to Kripke's *Naming and Necessity*.

[38] Example (26) is from Davies, pp. 172–3. Similar examples can be found in Richards, Wilson, Soames, and Ludlow and Neale.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Similarly where the antecedent is a definite description:

(28) *The inventor of the wheel* was a genius. I suspect *he/she* ate a lot of fish.

(29) I suspect that *the man who murdered Smith* was from out of town. In fact, I suspect *he* was a foreigner.

To the extent that it makes available *de dicto* readings, the proxy theory must therefore be viewed as superior to Evans's theory, at least for descriptive pronouns in contexts of propositional attitude. To avoid confusion, let us henceforth call pronouns that literally go proxy for definite descriptions D-type (rather than E-type) pronouns.[39]

What of descriptive anaphors in modal contexts? Are they E-type or D-type pronouns? Consider the following sort of sentence, discussed by Lauri Karttunen:[40]

(30) Mrs Jones wants Mary to marry *a rich man*. *He* must be a banker.

The *de dicto-de dicto* reading of this pair of sentences is perfectly natural. On that reading, the logical form of the antecedent sentence is given by:

(30_1) Mrs Jones wants ([an x: man x & rich x](Mary marry x)).

By (P4) the logical form of the antecedent clause is just:

(30_2) [an x: man x & rich x](Mary marry x).

If we now apply (P5) to 'he' in the anaphor sentence, that sentence as a whole comes out as either (30_3) or (30_4):

(30_3) [the x: man x & rich x & Mary marry x] ■ (banker x))

(30_4) ■ [the x: man x & rich x & Mary marry x](banker x))

[39] This terminology is borrowed from Fred Sommers, *The Logic of Natural Language* (New York: Oxford, 1982). Since it is something equivalent to the reading in which the descriptive pronoun has narrow scope that lies beyond Evans's grasp, it will not do to reply on his behalf that the ambiguities in (26)–(29) are due to the fact that the definite and indefinite descriptions they contain are ambiguous between Russellian and referential interpretations. Besides, (a) wide-scope and referential readings of descriptions are not the same thing, (b) descriptive pronouns may take intermediate scope in more complex cases, and (c) Evans himself favors unitary Russellian analyses of definite and indefinite descriptions.

[40] "Discourse Referents," in James McCawley, ed., *Syntax and Semantics*, VII: *Notes from the Linguistic Underground* (New York: Academic, 1976), pp. 363–385. The example Karttunen actually uses has 'Mary wants to marry a rich man' as its antecedent sentence. I have modified it so as to avoid distracting side issues concerning equi-NP deletion and *de se* attitude reports.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

where '■' stands for whatever modal operator is introduced by 'must'. And it is clear that it is (30_4) that is required where the antecedent sentence is read *de dicto*.[41]

I shall henceforth assume that all descriptive pronouns are D-type, i.e., that there are no E-type pronouns in natural language.

V. VARIABLE-BINDING AND RELATIVE CLAUSES

Let us return to the pronoun 'it' in (1):

(1) Every man who bought *a donkey* vaccinated *it*.

As we saw in section I, if the indefinite description 'a donkey' is treated as an existentially quantified noun phrase, then the anaphoric pronoun 'it' cannot be treated as a bound variable. The discussion in section IV provides an explanation of this fact: the indefinite description does not c-command the pronoun (the indefinite is buried inside a restrictive relative clause).

There are three questions before us here: (i) According to the theory of unbound anaphora developed in section IV, is the pronoun 'it' in (1) a genuine D-type anaphor? (ii) If so, does an application of (P5) to that pronoun result in an interpretable reading of (1)? (iii) If so, does the reading delivered by (P5) capture the (Geachian) truth conditions of (1)?

Clearly the answer to (i) is 'yes'. Since the pronoun in question is not c-commanded by its antecedent, by (P2) it cannot be bound, and by (P5) it must receive a descriptive interpretation. Intuitively, we want the sentence to be understood as something like:

(1_5) *Every man who bought a donkey* vaccinated the donkey *he* bought.

with 'he' bound by 'every man who bought a donkey'. And as we shall now see, this is exactly what a rote application of (P5) delivers.

In the logical notation we have been using, both determiners and quantifiers function as variable-binding operators. A determiner $\ulcorner Dx \urcorner$ binds any occurrence of x in the formula with which it combines to form a restricted quantifier. And a restricted quantifier

[41] The possibility of descriptive pronouns taking narrow scope appears to throw some light on Karttunen's observation that, if the antecedent sentence in (30) is given its *de dicto* reading, the modal auxiliary has to be there for the anaphora to work. Compare (30) with (i): (i) Mrs Jones wants Mary to marry *a rich man*. *He* is a banker. If the antecedent sentence in (i) is read *de re*, there is no problem involved in construing the pronoun in the anaphor sentence as anaphoric on 'a rich man'. But the anaphoric relation is problematic if the antecedent sentence is read *de dicto*. Unpacking the pronoun 'he' in accordance with (P5) yields only one reading for that clause because there is no modal operator with which the description may interact. And on this reading the description will give rise to an implication of existence that will be infelicitous (on Gricean grounds) if the antecedent clause is read *de dicto*.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

$\ulcorner[Dx\colon \phi]\urcorner$ binds any occurrence of x in the formula with which it combines to form a formula. We can think of relative pronouns that head restrictive relative clauses as variables bound by determiners. For example, we can interpret (31) as (32), and (33) as (34):

(31) Every man who loves Mary is happy.

(32) [every x: man x & x loves Mary](x is happy)

(33) Every man whom Mary loves is happy.

(34) [every x: man x & Mary loves x](x is happy)[42]

On the natural assumption that English determiners combine with complex nouns like 'man who loves Mary' and 'man whom Mary loves', in (31) and (33) the relative pronouns 'who' and 'whom' are c-commanded by the determiner 'every'. Consequently, if relative pronouns are treated as variables, we can specify a more general constraint on the operator-variable relationship in English (and perhaps other languages):

(P6) A variable v is bound by an operator Ov only if Ov c-commands v.[43]

On this account, the relative pronoun 'who' in our target sentence (1) functions as a variable bound by 'every', so the subject noun phrase 'every man who bought a donkey' has the logical structure given by (1_4):

(1) Every man who bought a donkey vaccinated it.

(1_4) [every x: man x & [a y: donkey y](x bought y)]

Now, what of the pronoun 'it' in the verb phrase? Since its antecedent, 'a donkey', does not c-command it, by (P2) the pronoun is not a bound anaphor. By (P5), it is therefore a D-type pronoun that goes proxy for a definite description recoverable from the antecedent clause. By (P4), the antecedent clause for the indefinite description 'a donkey' is:

(1_5) [a y: donkey y](x bought y)

[42] I have made a simplifying assumption here. Following Quine and Evans ["Pronouns, Quantifiers, and Relative Clauses (I)"], it is common to think of relative pronouns as devices of predicate abstraction rather than as genuine variables. On such an account, (31) and (33) are interpreted as (i) and (ii), respectively: (i) [the x: man x & [λz](z loves Mary)x](x is happy); (ii) [the x: man x & [λz](Mary loves z)x](x is happy). In what follows, however, it will do no semantical harm to treat relative pronouns as variables since (i) and (ii) are equivalent to (32) and (34), respectively.

[43] With minimal semantical and syntactical asumptions, (P2) follows from (P6).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

which contains a free occurrence of x. Applying (P5) to the anaphoric pronoun 'it', we get:

(1_6) [the y: donkey y & x bought y]

which represents the English description 'the donkey he bought'.[44] Thus, (1) as a whole will be interpreted as either (1_7) or (1_8), according as the subject quantifier or the descriptive pronoun is given wider scope:

(1_7) [every x: man x & [a y: donkey y](x bought y)]

([the y: donkey y & x bought y](x vaccinated y))

(1_8) [the y: donkey y & x bought y]

([every x: man x & [a y: donkey y](x bought y)](x vaccinated y))

(1_8) is no good because it contains a free occurrence of x. But (1_7) is formally impeccable. It represents the English sentence:

(1_3) *Every man who bought a donkey* vaccinated the donkey *he* bought.

on the reading on which the pronoun 'he' is interpreted as a variable bound by 'every man who bought a donkey' [the pronoun is c-commanded by the quantifier as (P2) requires].

There is no formal trick here. We have respected (P2) and (P6), used (P4) to determine the anaphor clause, applied (P5) in a rote fashion, and opted for the reading upon which the descriptive pronoun takes narrow scope.

As it stands, there is an obvious worry about the truth conditions of (1_3). But before examining this worry, let me pause to stress a formal point: the intelligibility of (1_7) refutes Evans's[45] claim that the anaphoric connection between 'a donkey' and 'it' in (1) precludes a treatment of the subject noun phrase 'every man who bought a donkey' as a restricted quantifier, i.e., as a "logical unit." The theory under consideration—a theory that (a) treats quantified noun phrases in English as restricted quantifiers, and (b) treats unbound

[44] A theory entertained by Terence Parsons (in "Pronouns as Paraphrases," manuscript, University of Massachusetts at Amherst, 1978) also cashes out the pronoun in terms of such a description.

[45] "Pronouns, Quantifiers, and Relative Clauses (I)," pp. 136–9; *The Varieties of Reference*, pp. 58–9. A point of clarification is in order here. Geach claims that 'man who bought a donkey', as it occurs in, e.g., 'every man who bought a donkey' is not a logical unit. Evans refutes this claim by constructing a viable semantical theory in which the phrase *is* a logical unit. But then Evans goes on to claim that 'every man who bought a donkey' as it occurs in, e.g., 'every man who bought a donkey vaccinated it' is not a logical unit. And Evans's claim has been refuted in the same way as Geach's, i.e., by the construction of a viable semantical theory in which 'every man who bought a donkey' *is* treated as a logical unit.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

pronouns that are anaphoric on quantifiers as definite descriptions —has delivered a perfectly intelligible reading of (1), viz., (1_7). Furthermore, (1_7) is equivalent to the reading that Evans claims his theory delivers for (1). So either both theories provide a correct account of the truth conditions of (1), or neither does: they stand or fall together.[46]

In my opinion, they fall (but not in a way that imperils the view that 'every man who bought a donkey' is a logical unit). It is not a consequence of (1) that every man who bought a donkey bought exactly one donkey; but this is a consequence of (1_7) because of the implication of uniqueness built into singular Russellian descriptions. Consequently, the D-type analysis given by (1_7) and the standard analysis given by (1_2):

(1_2) $(\forall x)(\forall y)((\text{man } x \ \& \ \text{donkey } y \ \& \ x \text{ bought } y) \supset x \text{ vaccinated } y)$

yield divergent truth conditions for (1). In particular, if there is a man who bought two (or more) donkeys and who vaccinated both (or all) of them, (1_7) will be false whereas (1_2) can still be true. I think it must be conceded that this a genuine failing on the part of the D-type analysis as it stands.[47]

A variation of the same problem arises with a D-type analysis of the pronoun in (2):

[46] Although Evans's claim is clearly false, there is no quick way to explain exactly where and why his argument breaks down; this is (mainly) because a charitable reconstruction of the argument turns crucially on certain technical details of Evans's own formal theories of quantification, anaphora, and relative clauses (cf. my "Binary Quantifiers and Unary Quantifier-formers," forthcoming). For present purposes, the following informal remarks will suffice. The breakdown of Evans's argument is connected in important ways to (a) his belief that descriptive pronouns cannot take narrow scope and (b) his preference for a "Fregean" interpretation of quantifiers. When certain constraints imposed by his own formal machinery are superimposed on these prior strictures, he is unable to countenance the idea that 'it' in (1) is equivalent to a description that takes narrow scope with respect to the quantifier 'every man who bought a donkey'. At most, Evans could hope to demonstrate that, if one endorsed his view that descriptive pronouns have their references fixed by description, and if one endorsed his "Fregean" interpretation of the quantifiers, then, in order to generate an interpretable reading of (1), one would have to treat 'every' as a binary quantifier instead of treating 'every man who bought a donkey' as a restricted quantifier. But we have already seen that there are cases in which descriptive pronouns take narrow scope. The upshot of all this is that, rather than provide support for the view that quantified noun phrases cannot be treated as restricted quantifiers, sentence (1) provides further support for the view that descriptive pronouns are D-type rather than E-type anaphors.

[47] This point is made by Heim, who provides a comprehensive discussion of the most notable attempts to deal with donkey anaphora before presenting her own theory. The same objection to Evans's theory is made by Richards before he presents his theory. Perhaps the same worry lies behind the failure of Kamp and Barwise to engage Evans's theory—Kamp makes only one fleeting reference to Evans's influential work and Barwise does not mention it.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(2) If John buys *a donkey* he vaccinates *it*.

Again, assuming that an indicative conditional like (2) is composed of two sentences and a binary sentential connective, the pronoun 'it' in (2) is a D-type rather than a bound anaphor because it is not c-commanded by its antecedent. A routine application of (P5) delivers:

(2_S) [An x: donkey x](John buys x) $\supset$

[The x: donkey x & John buys x](John vaccinates x)

which, unlike (2), cannot be true if John buys more than one donkey.

We appear to face a dilemma. The standard Geachian analyses of (1) and (2) provide plausible truth conditions at the expense of an unsatisfactory treatment of indefinite descriptions. The D-type analysis, on the other hand, respects the standard semantics of indefinite descriptions at the expense of unacceptable truth conditions for (1) and (2). In the next section, I shall look at a natural way of repairing the D-type analysis which does not compromise its general appeal.

VI. NUMBERLESS PRONOUNS

Typically, the syntactical number on a bound pronoun agrees with that of its antecedent.[48] For example, if we use 'all boys' we must use 'they', 'them', 'their', or 'themselves'; if we use 'every boy' we must use 'he', 'him', 'his', or 'himself':

(36) *All boys* respect *their* mothers.

(37) *Every boy* respects *his* mother.

In an introductory logic course, both (36) and (37) might be rendered as:

(38) $(\forall x)(Bx \supset Rx(\iota y)(Myx))$

which makes it clear that whether the pronoun is singular or plural is of no truth-conditional significance; number agreement between antecedent and anaphor is a matter of grammar. Of course, this is a quite general fact about bound anaphora. The variables of quantification theory are unmarked for number (and gender), so if a bound pronoun is to be treated as the natural language counterpart of the logician's bound variable, its overt number will be truth-conditionally inert.

With D-type anaphora matters are more complicated. In the examples we have looked at so far, anaphoric pronouns have agreed in syntactical number with their antecedents. And there has been an implicit assumption that the syntactical number of a D-type pronoun

[48] There are counterexamples, such as 'Someone has left their copy of the *Phaedo* behind', when uttered in an attempt to avoid using the masculine pronoun.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

determines its semantical number. Consequently, unbound pronouns anaphoric on ⌜just one F⌝, ⌜the F⌝, ⌜some F⌝, and ⌜an F⌝, have come out as singular definite descriptions, whereas those anaphoric on ⌜most Fs⌝, ⌜several Fs⌝, ⌜the Fs⌝, ⌜some Fs⌝, and ⌜a few Fs⌝, have come out as plural descriptions.

But what happens to unbound pronouns that are anaphoric on universally quantified noun phrases? Evans notes that there seems to be a problem with unbound anaphora on 'every'-phrases:

(39) *Every congressman* came to the party. *He* had a marvellous time.

This seems to be unacceptable on the anaphoric reading intended. Evans[49] offers an explanation in terms of a semantical number clash between antecedent and anaphor: (i) the pronoun is unbound (since its antecedent does not c-command it), therefore (ii) since it is singular, it is interpreted via the singular definite description 'the congressman who came to the party'; yet (iii) the antecedent sentence asserts the existence of a plurality of congressmen who came to the party.

There are problems with this explanation. First, it is not clear that the antecedent sentence in (39) is false if there is only one congressman. We know there are more, and this fact may be well affect the way we react to (39). Contrast (39) with (40):

(40) *Every frisbee major* got a job this year. *He* is very happy about it.

Of course, someone who uttered (40) would normally be aiming for some sort of rhetorical effect, but it is clear that the anaphora still works. A partial explanation of the joke-like quality of (40) surely lies in the fact that, if the speaker knows that exactly one person satisfies the descriptive condition expressed by 'frisbee major', typically she will use a definite description like 'the frisbee major' or 'our frisbee major', rather than 'every frisbee major'. It is not that 'every frisbee major' carries with it an implication of plurality, rather it is that it does not carry with it an implication of singularity. (As we shall soon see, the difference is important.)

The second problem with Evans's explanation is that it fails to explain the oddity of (39_1):

(39_1) *Every congressman* came to the party. *They* had a marvellous time.

Perhaps (39_1) is an improvement over (39), but it still seems a little strained. The reason, I suspect, is that in moving from (39) to (39_1) we have traded an alleged clash in semantical number for a very real clash in syntactical number.

[49] "Pronouns," p. 220.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

We are faced here with some quite general questions about number agreement and about cardinality implications present in maximal quantifiers. Recall that D-type anaphora is not always acceptable when the antecedent is of the form ⌜no F⌝ or ⌜no Fs⌝:

(41) John bought *no donkeys*. Harry vaccinated *them*.

(42) Every man who bought *no donkeys* vaccinated *them*.

In my opinion, however, it would be a mistake to construct a semantical theory that prevented unbound pronouns from being anaphoric on quantifiers of the form ⌜no F(s)⌝. First, an example like the following would be ruled out:

(43) Either there's *no bathroom* in this house or *it's* in a funny place.[50]

Second, the following sentences are just as bad as (41) and (42), but we do not want to prohibit unbound anaphora on indefinite descriptions:

(44) John didn't buy *a donkey* and Harry vaccinated *it*.

(45) Every man who didn't buy *a donkey* vaccinated *it*.[51]

The syntactical and semantical rules of the language should not conspire to block (41), (42), (44), and (45): they are perfectly well-formed. The problem is simply that, in the normal course of things, it would make no practical sense to use these sentences. The theory of D-type anaphora explains this. Consider (41); the anaphoric pronoun will come out as 'the donkeys John bought', so the sentence as a whole will be straightforwardly contradictory; similarly for the other examples.

Related considerations apply where the antecedent is a universally quantified noun phrase. In the light of the alleged anaphoric difficulties in (39) and (39_1), and in examples like (46) and (47):

(46) Every man who bought *every donkey* vaccinated *it*.

(47) Every man who bought *every donkey* vaccinated *them*.

some people seem to think that an adequate semantical theory must prevent pronouns from being anaphoric on 'every' phrases that do not c-command them.[52] In my opinion, this is a mistake. Consider (47). Since this sentence could be true only in a situation in which just one man bought any donkeys, it would certainly be much more natural to say:

[50] This example is due to Barbara Partee.

[51] Of course, (45) is fine if 'a donkey' is given wide scope, but we are not concerned that readings here (see fn. 6 and fn. 59).

[52] See, e.g., Hintikka and Carlson, Kamp, and Barwise.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(48) The man who bought *every donkey* vaccinated *them.*[53]

And notice that (48) falsifies the view that pronouns anaphoric on 'every' phrases must be c-commanded by them. So do the following:

(40) *Every frisbee major* got a job this year. *He* is very happy about it.

(49) Everyone who managed to listen to *every prophet* was asked to rank *them.*

(50) If John manages to acquire *every Rembrandt* he will build a museum in which to house *them.*

Let us now turn to the matter of semantical number. Even if it is true that phrases of the form ⌜every F⌝ and ⌜each F⌝ have existential import—which is itself open to debate—it is untrue that these quantifiers are always genuinely plural, and untrue that singular unbound pronouns can never be anaphoric on them. Consider the following sentences:

(51) *Every Swiss male over the age of twenty-one* owns a gun. *He* is required to do so by law.

(52) *Each candidate* will be debriefed by Mrs Hendrix. *He* will given some advice on how to tackle the press.

But as Evans notes, on the face of it, singular descriptive pronouns seem to carry implications of singularity, and plural descriptive pronouns seem to carry implications of plurality. The upshot of all this is that, in cases where we do not wish to prejudice the issue, we feel uneasy using either. Unbound anaphora on 'every' and 'each' phrases is awkward, however you look at it. To preserve syntactical agreement, a singular pronoun is required; but from a pragmatic standpoint, a plural pronoun will seem more appropriate unless there is doubt (or feigned doubt) as to whether there is more than one F. If I am certain that are several candidates for the new job, I might use (53):

(53) Mrs Hendrix introduced herself to *every candidate for the new job.* Later she had a reception for *them* over at her house.

But if I know that there is only one candidate, a certain John Jones, I will be inclined to use (53_1):

[53] There are conceivable circumstances in which one might also take utterances of (47) to be making stronger claims. Suppose that Mike bought every donkey on Monday, Tom bought them all from Mike on Tuesday, and Willy bought them all from Tom on Wednesday. The stronger reading I have in mind is the one that is true just in case Mike vaccinated every donkey, Tom vaccinated them, and Willy vaccinated them. (Notice the anaphora in this note.)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(53_1) Mrs Hendrix introduced herself to *the candidate for the new job*. Later she had a reception for *him* over at her house.

For rhetorical effect, however, in certain imaginable situations I might say:

(53_2) Mrs Hendrix introduced herself to *every candidate for the new job*. Later she had a reception for *him* over at her house.

But what about cases in which we do not want to commit ourselves one way or the other? One interesting idea is that in many such cases unbound anaphoric pronouns go proxy for definite descriptions that are, from a semantical perspective, numberless. The following sentences are surely equivalent:

(54) Every new recruit is armed.

(55) All new recruits are armed.[54]

Now suppose we want to continue with a sentence containing an anaphoric pronoun. Preserving syntactical agreement we get the following:

(54_1) *Every new recruit* is armed. *He* is ready for combat at a moment's notice.

(55_1) *All new recruits* are armed. *They* are ready for combat at a moment's notice.

Since the pronouns in these pairs are not bound by their antecedents, D-type analyses are called for. But this seems to present a problem. The antecedent sentences in (54_1) and (55_1) are surely truth-conditionally equivalent. But if the subsequent anaphoric pronouns are cashed out as singular and plural definite descriptions, respectively, the two anaphor sentences will not be equivalent. By (P5), the pronoun in (54_1) will come out as 'the new recruit', and the one in (55_1) will come out as 'the new recruits'. And on the analyses of singular and plural descriptions provided earlier, neither of these is right: (54_1) will now be false if there is more than one new recruit, and (55_1) will be false if there is exactly one new recruit. The anaphoric pronouns will have added cardinality implications not supplied by their antecedents. From the point of view of truth conditions, 'every new recruit' and 'all new recruits' are neither singular nor plural, which is why (54) and (55) are equivalent. Following up on a remark made by G. E. Moore, I suggested in section II that

[54] Similarly for 'Each new recruit is armed'. To say that these sentences are equivalent is not to say that 'every', 'all', and 'each' are "synonymous" in any strong sense. They may have rather different appropriateness conditions; since we are here concerned only with truth conditions, this can be ignored.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

phrases like 'whoever wrote *Waverley*' and 'whoever shot John F. Kennedy' might profitably be viewed as definite descriptions that are semantically numberless. Using ⌜[whe x: Fx]⌝ to represent a numberless (or neutral) description, the following truth clause was suggested:

(*8) ⌜[whe x: Fx](Gx)⌝ is true iff $|F - G| = 0$ and $|F| \geq 1$

With a view to honoring the truth-conditional neutrality of their antecedents, we might think of D-type pronouns that are anaphoric on quantifier phrases of the form ⌜every F⌝, ⌜all Fs⌝, and ⌜each F⌝ as semantically numberless.

VII. INDEFINITE DESCRIPTIONS AND NUMBERLESS ANAPHORS

If we take seriously the view that an anaphoric pronoun should not add cardinality implications not already supplied by its antecedent, we should expect to find numberless occurrences elsewhere. On Russell's account, a sentence of the grammatical form ⌜An F is G⌝ has the same logical form as a sentence of the form ⌜Some Fs are Gs⌝. In the notation of section II, ⌜[an x: Fx](Gx)⌝ and ⌜[some x: Fx](Gx)⌝ are logically equivalent; both are true if, and only if, there is at least one F that is G.[55] On a completely Russellian account of indefinite descriptions, then, since there can be no difference between the truth-conditional contributions of 'a donkey' and 'some donkeys', the same must be true of the complex noun phrases (i) 'every man who bought a donkey' and (ii) 'every man who bought some donkeys' (with minimal assumptions concerning compositionality). So in our target sentence (1), we should be able to substitute (ii) for (i) *salva veritate*. But notice that we cannot substitute *salva congruitate:* we must replace 'it' with 'them' to maintain syntactical agreement between anaphor and antecedent:

(1) Every man who bought *a donkey* vaccinated *it*.

(1_8) Every man who bought *some donkeys* vaccinated *them*.

Now, if we interpret singular and plural descriptive pronouns as singular and plural descriptions, respectively, not only do we get—as we have already seen—unacceptable truth conditions for (1), we get different truth conditions for (1_8). If 'them' is interpreted as 'the

[55] In *Introduction to Mathematical Philosophy,* Russell enters the following caveat: "rhetorically there is a difference, because in the one case there is a suggestion of singularity, and in the other case of plurality" (p. 171). Paul Grice's notion of conventional implicature might be of help in spelling out what Russell has in mind by "rhetorical effect" here. [See H. P. Grice, "Utterer's Meaning, Sentence Meaning, and Word Meaning," *Foundations of Language,* IV, 3 (1968): 225–242.] It is, of course, open to challenge Russell on the truth-conditional equivalence of ⌜An F is G⌝ and ⌜Some Fs are Gs⌝ (see below).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

donkeys he bought', (1_8) will be false if every man who bought a donkey bought just one donkey.[56] To the (complete) Russellian, this divergence might suggest that we pursue the suggestion, made in passing by Terence Parsons and by Martin Davies,[57] that descriptive pronouns anaphoric on indefinite descriptions are, or at least can be, interpreted as numberless descriptions. After all, independent evidence for the existence of such pronouns has already been provided. On such an account, the logical forms of (1) and (1_8) will be (1_9) and (1_{10}), respectively:

(1_9) [every x: man x & [a y: donkey y](x bought y)]

[whe y: donkey y & x bought y](x vaccinated y)

(1_{10}) [every x: man x & [some y: donkeys y](x bought y)]

[whe y: donkey y & x bought y](x vaccinated y)

On Russell's assumption that ⌜[a y: Fy](Gy)⌝ and ⌜[some y: Fy](Gy)⌝ are equivalent, (1_9) and (1_{10}) are also equivalent. Moreover, (1_9) gives us the right truth conditions for (1), viz., those given by Geach. So it does seem to be possible to provide an analysis of (1) that delivers the correct (Geachian) truth conditions and honors a Russellian treatment of singular indefinite descriptions. The "universalization" of the indefinite description 'a donkey' in (1) is a logical illusion: it is the pronoun that has universal force, by virtue of standing in for a definite description.

Let us look at some possible objections to this proposal. First, it might be objected that (1) and (1_8) have different truth conditions, that even if (1_9) captures the force of (1), (1_{10}) misrepresents the force of (1_8), the truth of which does not depend on how things are with any man who bought only one donkey. But this is not really an objection to the proposed treatment of the pronoun in (1_8), nor to Russell's treatment of singular indefinites; it is an objection to Russell's treatment of plural indefinites. The objector is contesting Russell's claim that ⌜An F is G⌝ and ⌜Some Fs are Gs⌝ are equivalent. In short, the objector is urging the following split:

(*4) ⌜An F is G⌝ is true iff $|F \cap G| \geq 1$

(*9) ⌜Some Fs are Gs⌝ is true iff $|F \cap G| > 1$

I am inclined to agree. But analyzing plural indefinites in this way

[56] Similarly for the following pair: (i) If John buys *a donkey* he vaccinates *it;* (ii) If John buys *some donkeys* he vaccinates *them.*

[57] See Parsons, p. 20, and Davies, p. 175. (I am grateful to Higginbotham for directing my attention to the former reference and to Davies for directing my attention to the latter.) Richards (p. 294) seems to be on the verge of making a similar suggestion.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

does not create a problem for the pronominal theory under consideration. If (*9) is adopted, the relevant difference in truth conditions between (1) and (1_8) will follow as a consequence of distinct treatments of the antecedents 'a donkey' and 'some donkeys' (rather than distinct treatments of the anaphors 'it' and 'them'). Second, it might be suggested that the numberless proposal cannot be correct because some pronouns anaphoric on singular indefinite descriptions are semantically singular. After all, in section IV we looked at some examples in which analyses of such pronouns as singular descriptions seemed to be appropriate.

There are several things one might say in response to this. First, one might maintain that, as a matter of fact, some D-type pronouns anaphoric on singular indefinites are singular while others are numberless. This would not, of course, be to abandon the spirit of the theory of D-type anaphora, it would simply be to bestow upon it a welcome degree of flexibility. The line of reasoning here would be as follows. When one uses a sentence of the form ⌜An F is G⌝, very often some singular belief or other furnishes the grounds for one's utterance, and a subsequent pronoun anaphoric on the indefinite will, in all probability, be interpreted as semantically singular. But if the indefinite description is within the scope of, say, a universal quantifier, very likely the speaker has general grounds for his assertion. In which case he is less likely to want to be committed to any implication of (relative) uniqueness. Consequently, if we are not to complicate the language by adding a new numberless pronoun, we must allow for the possibility of a numberless interpretation of 'it' in cases where the antecedent does not force a singular interpretation, i.e., where the antecedent is genuinely numberless.

And it is clear that singular indefinites themselves are genuinely numberless. Compare (*4) and (*5) above with (*6) and (*7):

(*6) ⌜The F is G⌝ is true iff $|F \cap G| = 1$ and $|F - G| = 0$

(*7) ⌜The Fs are Gs⌝ is true iff $|F \cap G| > 1$ and $|F - G| = 0$

The truth of ⌜the F is G⌝ requires that $|F \cap G| = 1$ and the truth of ⌜the Fs are Gs⌝ requires that $|F \cap G| > 1$, so these sentences can never be true together.[58] But singular indefinite descriptions are not semantically singular, they do not generate uniqueness implications. When we just wish to say that $|F \cap G| = 1$, we say ⌜Exactly one F is G⌝, or ⌜Just one F is G⌝. The truth of ⌜An F is G⌝ does not require that $|F \cap G| = 1$, it just requires that $|F \cap G| \geq 1$, and this is

[58] If $|F - G| = 0$, clearly $|F| = |F \cap G|$. I have used '$|F \cap G|$' rather than '$|F|$' in these restatements of (*6) and (*7) only to bring out the relationship between definite and indefinite descriptions.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

perfectly consistent with the truth of ⌜Some *F*s are *G*s⌝, even on the non-Russellian analysis of plural indefinites given in (*9). In short, singular indefinites are not semantically singular.

Now, according to the proposal under consideration, a pronoun anaphoric on a singular indefinite occurs in the singular so as to maintain syntactical agreement. At this stage we may proceed in one of two directions. One option, of course, is to say that the pronoun has to be interpreted as a numberless description so as not to add any cardinality implication to the truth conditions not already supplied by its antecedent. The other option is to say that the pronoun may be either singular or numberless depending upon various contextual or linguistic factors. In view of the fact that informants have insecure and divergent intuitions when questioned about the truth conditions and/or grammatically of many donkey sentences—see especially some of those listed below—I am inclined to think that the flexibility introduced by the latter option is called for. Compare the following:

(1) Every man who bought *a donkey* vaccinated *it*.

(57) Every man who has *a daughter* thinks *she* is the most beautiful girl in the world.

(58) Some man who bought *a donkey* vaccinated *it*.

(59) Some men who bought *a donkey* vaccinated *it*.

(60) Most men who bought *a donkey* vaccinated *it*.

As we have already seen, in (1) the numberless interpretation of the pronoun seems to be preferred. But as Parsons points out, in an example like (57)—which is of the same general form as (1)—a singular interpretation of the pronoun seems to be preferred. A reasonable explanation is that immediate linguistic context, and lexical and background knowledge, conspire to defeat the numberless interpretation (in the normal run of things, there cannot be two most beautiful girls in the world).

In (58) and (59), there is no obvious reason to prefer either interpretation, but the singular interpretation is apparently preferred; perhaps it is the default interpretation.[59] In (60) the numberless interpretation is preferred, probably for the same sorts of reasons it is preferred in (1).[60]

[59] The reading in which 'a donkey' takes wide scope and binds 'it' is also available for many speakers. [As mentioned earlier, sentence (1) also seems to admits of a reading in which 'a donkey' takes wide scope.] The existence of such readings seems to provide support for the view that there is a level of syntactical representation at which scope assignments have been made and at which (P2) holds.

[60] We can represent the logical form of (60) as: (i) [most x: men x & [a y: donkey y](x buys y)]([whe y: donkey y & x buys y](x vaccinates y)). This captures the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Sentence (61) provides very strong evidence for numberless D-type pronouns:

(61) Every man who bought *a beer* bought five others along with *it*.

If the pronoun 'it' is interpreted as a singular description then the

---

preferred truth conditions, on the assumption that (*5) is an adequate truth clause for 'most *F*s are *G*s': (*5) ⌜[most x: Fx](Gx)⌝ is true iff $|F \cap G| > |F - G|$. [Some informants get additional readings of (60), viz., one that requires that most men who bought at least one donkey vaccinated at least one of the donkeys they (qua individuals) bought, or another that requires that most men who bought at least one donkey vaccinated most of the donkeys they (qua individuals) bought.]

This is a good point to bring out the main differences between the theory of D-type anaphora and the theories of "discourse representation" (DR) proposed by Kamp and Heim, both of whom reject Russell's analyses of definite and indefinite descriptions as well as descriptive approaches to unbound anaphora. There are some differences between Kamp's theory and Heim's theory, but we can, I think, put these aside for present concerns. Two ideas stand out: (a) definite and indefinite descriptions introduce variables rather than quantifiers; (b) a quantified sentence ⌜[Dx: ϕ](ψ)⌝ is true if and only if D assignments of values to free variables in ϕ that satisfy ϕ also satisfy ψ. [As Heim stresses, this idea has its origins in one of the central proposals in David Lewis's "Adverbs of Quantification," in E. Keenan, ed., *Formal Semantics of Natural Language* (New York: Cambridge), pp. 3–15. The similarities show up especially in the analysis of conditional donkey sentences.] On this account, our target sentence (1) is true if and only if every assignment of values to x and y that satisfies (ii) also satisfies (iii): (ii) man x & donkey y & x bought y; (iii) x vaccinates y. Thus, the DR theory and the D-type theory (that allows number-neutral pronouns) deliver equivalent analyses of (1). But the ways in which these two theories achieve this result are, of course, rather different: the theory of D-type anaphora locates the implication of maximality in (1) in the anaphoric pronoun, whereas DR theory locates it in the determiner 'every'. Example (60), brings out this difference very clearly. On the DR proposal, (60) is true if and only if most assignments of values to x and y that satisfy (ii) also satisfy (iii). And as Richards (p. 281) points out, this is incorrect: suppose Alan bought ninety donkeys, and five other men bought exactly two donkeys each. DR theory predicts that (60) is true if Alan vaccinated fifty one or more of the donkeys he bought and the other five men failed to vaccinate any of their donkeys.

The problem here is that in (60) 'most' is quantifying over donkey-buying men and not over pairs of donkey-buying men and donkeys, as DR theory requires. This suggests that the success this approach has with capturing the implication of maximality in (1) is due to an artifact of first-order logic, and that it will need to be supplemented by some additional machinery if it is to handle the full range of examples. Whereas (60) seems to favor the D-type over the DR approach, it is arguable that (iv) provides evidence the other way (though this will ultimately depend upon one's final analyses of nonsingular descriptions and monotone decreasing quantifiers): (iv) No man who bought *a donkey* vaccinated *it*. On the DR proposal, (iv) is true just in case no man who bought a donkey vaccinated any of the donkeys he bought, which seems to me like the preferred reading. On the D-type proposal, if the pronoun goes proxy for the singular description 'the donkey he bought', we get a less desirable reading with an implication of relative uniqueness. If it goes proxy for a numberless description we get a reading equivalent to 'No man who bought a donkey vaccinated each of the donkeys he bought', which does not seem like a genuine reading of (iv). As Irene Heim has pointed out to me, (iv) does not actually pose a problem for the view that the pronoun goes proxy for a numberless description—on its preferred reading, (iv) can be paraphrased as 'No man who

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

sentence will be automatically false.[61] But if it is interpreted as a numberless description, things come out exactly right: the sentence will be true just in case every man who bought at least one beer bought five other beers along with each of the beers he bought.[62]

Further evidence for semantically numberless pronouns comes from sentences like the following:

(62) Every man who bought *a donkey and a mule* vaccinated *them*.

(63) Every man who bought *a donkey or a mule* vaccinated *it*.

---

bought a donkey vaccinated the donkey or donkeys he bought'—rather it poses a more general problem concerning the interpretation of nonsingular definite descriptions within the scope of monotone decreasing quantifiers. The following examples bear out Heim's point: (v) *No man* vaccinated the donkeys *he* bought; (vi) *Few men* vaccinated the donkeys *they* bought. Again, the descriptions are naturally interpreted as 'any of the donkeys he (they) bought' rather than 'each of the donkeys he (they) bought'.

The situation, then, seems to be as follows: (a) DR theory has problems with a sentence like (60), in which the indefinite description is within the scope of a non-first-order determiner such as 'most'; (b) if the D-type theory utilizes the standard distributive analysis of plural and numberless descriptions, then it will have a problem with a sentence like (iv), in which the indefinite description is within the scope of a monotone decreasing determiner like 'no'. Consequently, if we use a determiner like 'few', which is both monotone decreasing and non-first-order, we can construct a sentence that may be problematic for both approaches: (vii) Few men who bought *a donkey* vaccinated *it*. A further difference between the D-type and DR approaches is that the latter, along with the theories proposed by Hintikka and Carlson and by Barwise, explicitly prohibits the anaphoric relations in sentences where the antecedent is of the form ⌜every F⌝—e.g., (48)–(52)—because 'every' phrases are genuine quantifiers (rather than referring expressions or devices that simply introduce variables).

In this note, I am indebted to discussion with Irene Heim.

[61] The problem for Evans posed by sentences like (61) was pointed out by Heim and by Ernest Le Pore and James Garson in "Pronouns and Quantifier-Scope in English," *Journal of Philosophical Logic*, XII, 4 (1983): 327–358.

[62] A numberless analysis of the pronoun 'it' in (61) does not mean that the sentence cannot be true unless every man who bought a beer bought eleven beers; nor does it mean that the sentence cannot be true unless every man who bought a beer bought an infinite number of beers (in discussion periods, it has more than once been claimed that the numberless analysis makes one or other of these undesirable predictions). As I mentioned earlier, I am confining my talk about plural and numberless descriptions to cases in which they receive distributive readings. This is reflected in the truth clauses (*7) and (*8). It is well-known that many plural noun phrases may be interpreted collectively when combined with the right predicates; and, naturally enough, this includes plural descriptions ('The donkeys pulled Pedro's cart up the hill') and plural descriptive pronouns ('*The men* hauled Bill's Piano down the stairs; then *they* loaded it into a truck'). It is, of course, the simple distributive reading of the numberless descriptive pronoun 'it' in (61) that we are interested in. Confused talk of men buying eleven (or an infinite number of) beers results from thinking that nonsingular descriptive pronouns have to be interpreted collectively even though nonsingular descriptions do not. I suspect that the origin of this confusion lies in popular but loose talk of plural E-type pronouns as referring to "sets," "classes," "maximal collections," and the like. Evans himself says nothing that commits him to just collective readings of nonsingular E-type pronouns, and for good reason.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(64) Every man who bought *more than one donkey* vaccinated *it*.

(65) Every man who bought *two or more donkeys* vaccinated *them*.

In (62), the pronoun appears to be going proxy for the conjunctive description 'the donkey(s) he bought and the mule(s) he bought', which is syntactically plural as well as semantically plural (given the semantics of conjunction), so 'them' is the correct choice. In (63), the pronoun appears to be going proxy for 'the donkey(s) he bought or the mule(s) he bought'. The semantics of disjunction requires that, for (63) to be true, every man who bought a donkey vaccinated it, every man who bought a mule vaccinated it, and every man who bought a mule and a donkey vaccinated them; hence the pronoun seems to demand a numberless interpretation. (This suggests, perhaps, that when a numberless interpretation is required and there is no syntactical reason to use a plural pronoun, it is the singular pronoun that is called upon.)

Many speakers find (64) and (65) truth-conditionally equivalent. But notice that in (64), the pronoun is syntactically singular so as to agree with 'more than one donkey', even though it cannot be interpreted as semantically singular for fear of contradiction.

It seems to be clear, then, that donkey anaphora of the sort exemplified in sentences of this general form—i.e., where the antecedent is a constituent of a relative clause—does not by itself undermine either a unitary Russellian analysis of indefinite descriptions or a D-type analysis of unbound anaphora. The maximality implication that (1) apparently gives rise to can be captured once it is seen that D-type pronouns anaphoric on numberless antecedents may, and in some cases must, receive numberless interpretations, and that agreement between anaphors and antecedent noun phrases is, by and large, a syntactical matter.

Virtually everything that has just been said carries over *mutatis mutandis* to conditional donkey sentences. Notoriously, conditionals give rise to all sorts of problems in philosophical logic, and it is with great trepidation that I bring them up at all. It is worth remarking, however, that, under the present proposal, anaphoric relations in conditionals do not appear to raise any new or additional problems. Indeed, the proposal seems to help with some old ones. Consider (2):

(2) If John buys *a donkey* he vaccinates *it*.

The truth conditions of (2) are apparently captured by (2_3), a fact that allegedly commits us to a wide-scope universal rather than a narrow-scope existential treatment of the indefinite description:

(2_3) [Every x: donkey x](John buys $x \supset$ John vaccinates x)

But we can now provide a coherent account of (2) that comports with the existential analysis of the indefinite. On the assumption that a

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

conditional is composed of two sentences and a binary sentential connective, the pronoun 'it' in (2) is not a bound anaphor because it is not c-commanded by its antecedent. So on the numberless D-type proposal, (2) will come out as (2_4):

(2_4) [an x: donkey x](John buys x) $\supset$

[whe x: donkey x & John buys x](John vaccinates x)

which does not seem to invite any objection (over and above those standardly brought up against the material conditional). It would seem, then, that we are not forced to treat indefinite descriptions in the antecedents of conditionals as universally quantified expressions that mysteriously take wide scope.

This is a good point at which to say something about quantifiers of the form ⌜any F⌝. It is sometimes claimed that quantifiers of this form are special in that they appear to insist on wide scope in certain environments.[63] In particular, it is sometimes argued that the scope of an occurrence of ⌜any F⌝ in the antecedent of a conditional extends to the consequent. Consider the following example (from Chrysippus):

(66) If *any man* was born at the rising of the dog-star, *he* will not die at sea.[64]

By treating 'any man' as a universally quantified expression with wide scope over the entire conditional, and by treating 'he' as a bound variable, one can capture the force of (66) with (66_1):

(66_1) [Every x: man x]

(x was born at the rising of the dog-star $\supset$ x will not die at sea)

But this is thoroughly unilluminating; it forces us to attribute to ⌜any F⌝ properties that other quantifiers simply do not possess, viz., the capacity to take scope over a conditional when embedded in its antecedent.

Under the D-type proposal, we no longer need to treat 'any' as exceptional. On the assumption that (a) it receives its existential reading in (66), and (b) its scope is restricted to the antecedent of the conditional (i.e., everything it c-commands), the pronoun 'he' is a D-type anaphor and comes out as a numberless description, which gives us the following:

(66_2) [any x: man x](x was born at the rising of the dog-star) $\supset$

[whe x: man x & x was born at the rising of the dog-star](x will not die at sea)

[63] See, e.g., Hornstein. Evans ["Pronouns, Quantifiers, and Relative Clauses (I)"] also expresses some sympathy for this view.

[64] The example is taken from Cicero, *De Fato*, VI, in *Libri De Divinatione et De Fato* (Cambridge: Knapton, Kinnerock, & Vaillant, 1721), pp. 311–2.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

I should finally say something about how the present proposal handles the following example, which has been regarded as presenting insurmountable problems for descriptive approaches to unbound anaphora:

(67) If a man shares an apartment with another man, he also shares the housework with him.[65]

If the pronoun 'he' goes proxy for a singular description, the nature of the problem is clear: there is no unique man who shares an apartment with another man. For suppose b is such a man; then there is a distinct man c with whom b shares an apartment. But *sharing an apartment with* is symmetric, therefore c is also a man who shares an apartment with another man, viz., b.[66]

But now suppose the pronoun 'he' goes proxy for a numberless description. In order that we do not get swamped with quantifiers, let us use P as shorthand for the two-place predicate 'shares an apartment with', and R as shorthand for the two-place predicate 'shares the housework with'. The logical form of the antecedent of the conditional (67) is given by:

(67_1) [an x: man x]([a y: man y & $y \neq x$](Pxy))

By (P4) this also represents the antecedent clause for 'a man'. If we apply (P5) to the pronoun 'he' (in the consequent) we get:

(67_2) [whe x: man x & [a y: man y & $y \neq x$](Pxy)]

By (P4) the antecedent clause for 'another man' is:

(67_3) [a y: man y & $y \neq x$](Pxy)

If we apply (P5) to the pronoun 'him' we get:

(67_4) [whe y: man y & $y \neq x$ & Pxy]

Combining these descriptions with the matrix clause Rxy, the consequent of (67) comes out as:

(67_5) [whe x: man x & [a y: man y & $y \neq x$](Pxy)]
([whe y: man y & $y \neq x$ & Pxy](Rxy))

which says that every man who shares an apartment with another man shares the housework with every other man with whom he shares an apartment; and this is exactly what we want. Sentence (67) would therefore seem to provide very strong evidence for the view

[65] This particular example is due to Heim.
[66] The special feature of this example is that it resists attempts to secure uniqueness by appealing to an implicit event (or situation) parameter: there is no event (or situation) in which a unique man shares an apartment with another man.

CC-0. In Public Domain. Panini Kanya Maha Vidyalaya Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that D-type pronouns anaphoric on singular indefinite descriptions can be numberless.[67]

VIII. CONCLUSION

I have attempted to defend the view that the existence of donkey anaphora—indeed, unbound anaphora more generally—does not force us to abandon traditional accounts of quantification and anaphora. As Evans has demonstrated, pronouns anaphoric on quantifiers must stand in a particular syntactical relationship to those quantifiers if they are to be interpreted as bound variables. Those which do not stand in this relationship—essentially, those which are not constituents of the smallest clauses containing their antecedents—are plausibly interpreted as definite descriptions.

I have advocated three departures from Evans's theory. (i) With a view to preserving the syntactical and semantical unity of quantified noun phrases, I have treated such phrases as restricted quantifiers (rather than treat determiners as binary quantifiers). (ii) On Evans's account, descriptive pronouns are rigid designators whose referents are fixed by description. But there is clear evidence that descriptive pronouns may enter into scope interactions in much the same way as overt descriptions; consequently, I have treated such pronouns as going proxy for descriptions.[68] On such an account, the donkey pronoun in (1) goes proxy for a description that may take narrow scope with respect to the subject quantifier. The syntactical and semantical details of the theory now allow us to treat both the subject quantifier and the descriptive pronoun as restricted quantifiers, contrary to Evans's claims. (iii) Since certain quantifier phrases (including many universally quantified noun phrases and singular indefinite descriptions) are truth-conditionally numberless, it is natural to allow for the possibility that (at least some) occurrences of pronouns anaphoric on such quantifiers respect this number neutrality, agreement between anaphors and antecedents being essentially a syntactical matter. Sentence (1) is a case in point.

I hope I have, at least for now, made a plausible case for the view that the examples considered in this paper do not undermine (a) a unitary existential analysis of indefinite descriptions, (b) an account of quantified noun phrases as restricted quantifiers, or (c) a descriptive analysis of unbound anaphora.

STEPHEN NEALE

Princeton University

[67] Examples involving two or more D-type pronouns bring up other complex issues that are discussed in my *Descriptions*, chs. 5–6.

[68] A residual question here is whether there is a level of syntactical description at which the description appears.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

COMMENTS AND CRITICISM

THE POVERTY OF PLURALISM: A REPLY TO STERELNY AND KITCHER

In *The Nature of Selection,*[1] I argued that "genic selectionism" is an ambiguous thesis. The view of natural selection that this label denotes was developed by George Williams in his landmark book *Adaptation and Natural Selection.*[2] It then was popularized by Richard Dawkins in *The Selfish Gene;*[3] later, Dawkins elaborated and somewhat reformulated the idea in *The Extended Phenotype.*[4]

In saying that genic selectionism is ambiguous, I meant that the thesis has a trivial and a substantive formulation. I claimed that Williams and Dawkins confused these; they repeatedly conclude that the substantive thesis is true when all their arguments demonstrate is that the trivial version is correct.

The substantive thesis might roughly be expressed by saying that all adaptations exist solely because they benefit individual genes (or genotypes). The trivial thesis is the idea that all natural selection can be "represented" in terms of genes and their properties. Both these formulations require clarification, which I tried to provide in my book.

In a recent article, Kim Sterelny and Philip Kitcher[5] defend a version of genic selectionism and attempt to refute the criticisms I made of that doctrine. Their defense has two components. First, they find fault with the account I gave of the units-of-selection controversy—an account which uses the idea of probabilistic causality as a tool of explication. Second, they provide a positive account of their own of what that controversy concerns, one which they think allows genic selectionism to emerge as a successful thesis.[6]

[1] Cambridge: MIT, 1984. Hereafter NS.
[2] Princeton: University Press, 1966.
[3] New York: Oxford, 1976.
[4] San Francisco: Freeman, 1982.
[5] "The Return of the Gene," this JOURNAL, LXXXV, 7 (July 1988): 339–361. Cf. also Kitcher, Sterelny, and Waters, "The Illusory Riches of Sober's Monism," this JOURNAL, this issue, 158–161.
[6] The principal claims advanced by Sterelny and Kitcher are also developed by Ken Waters, "Environments, Pragmatics, and Genic Selection" (address to the American Philosophical Association, Eastern Division, 1985) and *Models of Natural Selection: From Darwin to Dawkins* (doctoral dissertation, Indiana University, 1986); see also his "Tempered Realism about the Forces of Selection" (forthcoming). John Cassidy defends a conventionalist thesis about the units-of-selection controversy in "Philosophical Aspects of the Group Selection Controversy," *Philosophy of Science,* XLV (1978): 575–594.

0022-362X/90/8703/151–8
© 1990 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

I believe that the position they sketch is mistaken, both in its general orientation and in its details. I believe that the Sterelny/Kitcher position misunderstands what the biological question of the units of selection is about and that their criticisms of my own proposal are mistaken as well. Due to limitations of space, I shall focus on the positive proposal that Sterelny and Kitcher advance.

The idea of genic selectionism arose as an answer to a question. The question was an old one, predating (though surviving into) the modern synthesis' integration of Mendelian genetics and the theory of natural selection. The general question would now be formulated by asking what the "units of selection" are. But the pressing special case of this question—the one that drove the problem from the time of Darwin and Wallace at least until Williams's 1966 book—was whether adaptations evolve in nature because they are good for the group (or species) in which they occur. Williams's book delivered an emphatic *no* to this question about group adaptation. Group adaptations require the process of group selection.[7] Williams argued that virtually no solid evidence exists for group selection and that it is more parsimonious to hold that group adaptation is rare, if not nonexistent.[8] But rather than return to the traditional idea that adaptations exist for the good of the organisms possessing them, Williams opposed the idea of group adaptation by invoking a new alternative. This was the idea that the gene—not the group or the individual organism—is the unit of selection. Genic selectionism was born as an alternative to the idea of group selection.

To defend this new thesis, one must do more than reject the ideas of group selection and group adaptation. It was at this point that the fundamental ambiguity in the genic-selectionist position emerged. Williams and Dawkins placed heavy emphasis on the idea that all natural selection processes can be "represented" in terms of the fitness values of individual alleles (or genotypes).[9] Somehow this fact was supposed to support a substantive thesis. What they did not

[7] Roughly speaking, an adaptation is a trait that exists because there was a selection process of a certain sort. A trait is a group adaptation, as opposed to simply providing a fortuituous group benefit, only if it evolved because it conferred a group benefit.

[8] Although Sterelny and Kitcher say that discussion of parsimony in connection with the units-of-selection issue is "unnecessary dust" kicked up by philosophers (340), I believe that credit should be given where credit is due. Parsimony and a variety of other philosophically loaded concepts are repeatedly pressed into service by Williams and Dawkins. Such considerations were not invented by card-carrying philosophers, but were put into play by the scientists themselves.

[9] An explication of what it means to "represent" a selection process in the vocabulary of genes and gene frequencies, and an explanation of why this fact about representability cuts no ice in the units-of-selection controversy, is provided in NS, ch. 7.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

realize is that the point about representation is a truism; it would be correct even if group selection and group adaptation were ubiquitous.

It is an empirical question whether there are group adaptations. Properly formulated questions about group selection require detailed information about population structure to be answered. More generally, I would claim that substantive biological questions about the units of selection have always been empirical. Hence, a litmus test for any philosophical clarification of this scientific problem is that it clearly identify what the empirical issues really are.

Sterelny and Kitcher advocate a pluralistic view of this controversy. According to them, there are always "alternative, maximally adequate representations of the causal structure of the selection process" (358). I believe that there is a sense in which they are right—but it is trivial and utterly disconnected from the biological problem. To illustrate what I have in mind here, I shall describe a case of group selection and group adaptation. Then I shall describe what Sterelny and Kitcher's pluralism would say about the case in question.

Just as the focal point of the units-of-selection controversy has been the issue of group selection, so the focal point of the group-selection controversy has been the issue of altruism. An altruistic characteristic[10] of an organism is one that hurts the individual that has it while helping the group within which the individual lives. Help and hurt are calculated in the currency of survival and reproductive success—of fitness. The opposite trait—"selfishness"—leads its bearer to enjoy the donations provided by altruists, even though selfish individuals do not reciprocate. Selfish individuals are free riders. In a single population made of altruists and selfish individuals, altruists will be less fit than selfish individuals, and so (with the usual assumptions about heritability, etc.) the population will evolve to a state in which altruism has disappeared.

The following graph of the fitnesses of altruists (A) and selfish individuals (S) illustrates two fundamental facts about these characteristics. No matter what the mix is within a group of As and Ss, the A individuals are on average less fit than the S individuals. The fitness of a group however (as measured by the average fitness of the organisms in it, $\bar{w}$), is enhanced by containing high concentrations of altruists.

[10] In the evolutionary, not the psychological, sense of that term. I discuss the difference in "What is Evolutionary Altruism?", B. Linsky and M. Matthen, eds., *New Essays on Philosophy and Biology* (*Canadian Journal of Philosophy* Supplementary Volume), XIV (1988): 75–99; and in "What is Psychological Egoism?" *Behaviorism* (forthcoming).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

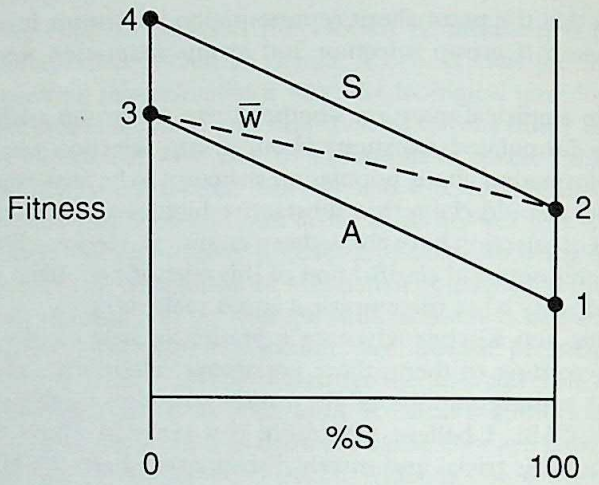


Figure 1

I said before that, if the two traits evolve in the context of a single cohesive population, *S* will displace *A*. This need not be true, however, if a process of group selection is underway. Imagine that we have an ensemble of groups, each containing some mix of *A* and *S*. If certain contingent conditions are satisfied, *A* can evolve and be maintained in the ensemble.

The following two biological conditions roughly suffice. First, let like live with like. Second, let the populations fragment and found new colonies fast enough (i.e., before the process of individual selection at work within each group drives *S* to fixation). Even though I have stated these biological requirements in only a rough and qualitative way, the main conceptual difficulty in understanding how altruism can evolve under group selection may now be stated: If *S* is fitter than *A* within each group (a definitional truth), how can *A* evolve? Within each group, *S* is declining in frequency. Does this not mean that *S* must disappear from the ensemble of groups?

A beginning of an answer comes from the requirement that like live with like. Imagine that the ensemble consists of 200 organisms, which live in two groups of 100 individuals each. The first group is 99% *S* and the second is 1% *S*.[11] Below are the average fitnesses (the *w*s) of selfish and altruistic individuals within each group, and the overall average fitnesses of the two traits. I have used the numbers on

[11] Note that the global frequencies of *A* and *S* are each 0.5; yet, most altruists live in a group in which the frequency of altruism is far greater than 50%. This is part of what I mean by 'like living with like'; there is no requirement that the groups be internally homogeneous.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the y-axis in the above figure to represent fitnesses (think of them as expected numbers of offspring if you like, or as arbitrary fitness units):

Group 1	Group 2	Overall Average
(99) $w_{S1} = 2$	(1) $w_{S2} = 4$	(100) $w_S \approx 2$
(1) $w_{A1} = 1$	(99) $w_{A2} = 3$	(100) $w_A \approx 3$

This is an example of Simpson's paradox;[12] inequalities true within each group reverse when one averages over all groups. Within each group, selfish individuals are on average fitter than altruists, but (in this example), the reverse is true when one considers the two population ensemble as a whole.[13]

For altruism to increase in frequency under group selection, altruists must be fitter than selfish individuals in the global ensemble. The fact that altruists are less fit within each group does not rule out that possibility. Rather, if like live with like, Simpson's paradox can take hold, and the altruistic trait can increase in frequency.

If the kind of process just sketched led a genuinely altruistic characteristic to evolve, a biologist would rightly see group selection at work. On the other hand, if selection occurs exclusively within the confines of a single cohesive population, then biologists would be right to view group selection as a fantasy. And if natural populations sometimes are one way and sometimes the other, then different answers about the units of selection would be appropriate in different cases. It all depends on the biology.

This may sound like pluralism, but it is not the pluralism that Sterelny and Kitcher espouse. I claim that, if an altruistic character evolves in the kind of circumstances just described, then one must interpret this as a case of group selection; Kitcher and Sterelny claim that all selection processes, including this one, can be described in a number of equally correct ways. Even if the process I have just described occurred in nature, they would say that there are many correct descriptions of it, ones which generate different answers to the question of what the units of selection are.

To show what this means, I shall add a further detail to the example. Let *A* and *S* be traits in a plant species in which an individual may manufacture an insecticide at some energetic cost to itself and then

[12] For philosophical discussion of the meaning of Simpson's paradox and its importance for understanding causality, see Nancy Cartwright, "Causal Laws and Effective Strategies," *Nous*, XIII (1979): 419–437.

[13] This example does not illustrate how *A* can increase in frequency after it is introduced by a single mutant or migrant. Rather, the point is to show how *A* can increase in frequency even though there is individual selection against it within each group.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

leach the chemical into the soil. Imagine that the insecticide protects not just the plant that produces it, but the plants that live near the producer as well.[14] If so, producing the insecticide and leaching it into the soil would be an altruistic trait. The selfish alternative would be the trait of not producing the insecticide. I am supposing that *S* individuals receive the benefit of the insecticide without incurring the energetic cost of producing it.

So *S* and *A* are phenotypic characteristics. What about the genetics? Suppose the organisms in question are haploid and that the two phenotypes are controlled by a single gene. Individuals with the *a* allele (at a given locus) produce the insecticide; individuals with the *s* allele do not. Under group selection, the *A* phenotype evolves; and so, of course, does the *a* gene.

Kitcher and Sterelny think that the following three statements are true of the process I have described:

(i) Groups that contain high frequencies of altruists do better than groups that do not; the reason is that containing lots of altruists is good for the group.
(ii) Altruistic individuals on average do better than selfish individuals. The reason is that altruists tend to live together, so they usually do not get exploited by free riders.
(iii) The *a* gene on average does better than the *s* gene. The reason is that *a* genes tend to occur in groups in which there are mainly other *a* genes, so *a* genes usually do not get exploited by free-rider genes.

I agree that these are all true, but it seems to me that nothing interesting follows from that. If Kitcher and Sterelny's pluralism is simply the idea that there are different vocabularies (group, organismic, and genic) that can be used to describe what is going on here, then the thesis is trivial. What is biologically significant is the process, not how it is described. Genes cause organismic phenotypes, and the phenotypes of organisms determine the phenotypes of the groups in which they live. This much is undeniable. But it does not follow that the above process can be reinterpreted so that one no longer needs to think of it as involving group selection.

Here is another way to see why Sterelny and Kitcher's pluralism does not engage the biological issue. To address the units-of-selection controversy, one must say which objects in nature show significant degrees of functional organization and explain why this is so. Organisms typically show a very high degree of functional organization. Do groups? Arguments about what the units of selection are

[14] For simplicity, suppose that a producer and the nonproducers that are its neighbors are protected to the same degree.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

attempt to answer this question and explain why the answer is *yes* or *no*. I see no room for conventionalism here.[15]

Organisms can be viewed as groups of integrated genes. An organism is a group; individual genes are its parts. Just as the group-selection controversy shows that there can be conflict between what is good for the group and what is good for the individual organism, it also should be possible for there to be conflict between what is good for the organism and what is good for the individual gene.[16] In the group-selection case, it is a matter of contingent biological fact that group adaptations are rare; if the interests of organisms usually win out over the interests of single genes, that, too, will be a contingent biological fact, one that calls for explanation. Looked at in this way, the thesis of genic selectionism is substantive—as vulnerable to empirical test as the thesis of group selectionism. A formulation of it that turns it into a truism has deprived it of its biological interest.

Kitcher and Sterelny argue that there are certain sorts of selection processes that are not consistent with the idea that the individual organism is the unit of selection. Here they make a point that I had developed in my book—that meiotic drive involves selection at the genic, not the organismic, level. When it comes to altruism, they think that this, too, is a phenomenon that cannot be handled organismically. And, of course, it goes without saying that for them there are some characters that cannot be viewed as group adaptations. So, for them, the thesis that the organism is the unit of selection is vulnerable to empirical disconfirmation; it is true sometimes, but not always. And the idea that the group is the unit of selection has the same empirical status. Yet, curiously enough, their view is that the genic perspective can accomodate anything and everything that might happen in natural selection. I, on the other hand, think that the standards applied to test whether groups or organisms are units of selection also should be applied to genes. When this is done, a monolithic genic selectionism is no more empirically plausible than a monolithic thesis of organismic or of group selectionism.

Sterelny and Kitcher share my desire to distinguish trivial genic selectionism from its substantive alternative. They agree that it is trivial that "the result of selection is (almost always) an increase in

[15] Unless one is prepared to be a conventionalist about causation, or about natural history, or about something far more global than the biological questions themselves. Sterelny and Kitcher do not seem inclined to take that radical route.

[16] For an interesting argument to the effect that fundamental facts about ontogeny are explained by the conflict of selection within organisms and selection between them, see Leo Buss, *The Evolution of Individuality* (Princeton: University Press, 1987). See also David Sloan Wilson and Elliott Sober, "Reviving the Superorganism," *Journal of Theoretical Biology*, CXXXVI (1989): 337–356.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

frequency of some gene in the gene pool" (339). They claim that Dawkins advances a second thesis, however, one that is supposed to be biologically important to the units-of-selection problem; this is the idea that "barring complications, the average ability of the genes in the gene pool to leave copies of themselves increases with time" (340). My view is that this second thesis is more or less beside the point.

When group selection manages to increase the global frequency of the *a* gene, it will be true that the average fitness of genes in the global population increases. The average fitness of the alleles in the population (where these are perfectly correlated with their respective phenotypes) is depicted in the above figure by $\bar{w}$. On the other hand, when purely individual or genic selection within a single cohesive population eliminates both the *A* phenotype and the *a* allele, $\bar{w}$ will decline. Group selection can lead the average fitness of genes to increase and genic selection can lead the average fitness of genes to decline. How often this quantity goes up or down has little to do with what the units of selection are.[17]

In my book, I claimed that it often is useful to represent selection processes in a genic vocabulary. This technique has generated some confusions; it also has provided a number of important insights. But the utility of a mode of representation is not what the units-of-selection problem has been about. It is about the levels at which adaptations have accumulated. The pluralism of Sterelny and Kitcher confuses these two questions, in much the same way that Williams and Dawkins did earlier.

ELLIOTT SOBER

University of Wisconsin/Madison

THE ILLUSORY RICHES OF SOBER'S MONISM

Biologists have discovered that a wide variety of population-genetics models apply to cases of natural selection in the wild. The outcome of selection may depend on the number

[17] Also, I am not aware that Dawkins has ever taken a stand on whether the average fitness of genes in a population typically increases. In *The Extended Phenotype* (p. 193), he mentions in passing that "many population geneticists are, for reasons best known to themselves, very interested in another quantity which is called the mean fitness of a population."

0022-362X/90/8703/158–161 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of loci, linkage among the loci, the number of alleles at a locus, whether or not selection is frequency-dependent, the relations among fitness parameters, and so forth. Some biologists want to make a further distinction among types of selection. They think that selection sometimes acts on the gene, sometimes on the organism, and sometimes on the group. They assume that, for each selection episode, there is a unique account that will identify the level of selection. Elliott Sober[1] hopes to legitimate this discourse and to explicate it in a way that will enable the biologists who like it to settle their differences empirically. We believe that asking about the real unit of selection is an exercise in muddled metaphysics.

Sober outlines an example that is intended to embarrass pluralism. Imagine a haploid plant species in which bearers of the *A* allele manufacture and leach into the soil a substance that benefits them and their neighbors, while bearers of the *S* allele do not. If one assumes that like plants associate, then a little algebra reveals that the *A* allele can be favored by selection. Imagine that Sober's example were expanded to a genuine model in which conditions were imposed to require that the associations would be stable across generations. Then we would indeed have a case that could be described from any of three perspectives: there would be group-level, organism-level, and genic-level descriptions in the manner of his statements (i)–(iii) (*ibid.*, p. 156). Biologists, we imagine, would agree on the algebra. We claim that there is no point in fighting about the "correct" interpretation. None of the descriptions should be singled out as the correct analysis that exposes the "true unit of selection."

Sober is opposed to this ecumenical conclusion because he thinks that the discovery of an example of this type would vindicate those who have been contending for group selection. It would. For such a discovery would show something not known in advance, namely, that some episodes can be adequately described by supposing that selective forces act on groups. That would be a real empirical discovery. Whether it would vindicate traditional Wynne-Edwards style group selection is another matter. After all, this style of evolutionary analysis was invented by their opponents.

Still, the empirical discovery of a process that can be described as group selection does not invalidate individual-level or gene-level descriptions. It does not begin to establish that there is a uniquely correct description of the causal process. The discovery that we can describe the visual system as applying an algorithm to infer shape

[1] See "The Poverty of Pluralism: A Reply to Sterelny and Kitcher," this JOURNAL, this issue, 151–8.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

from motion does not demonstrate that we cannot also describe it in terms of interacting neural assemblies.

Many events can be analyzed from more than one perspective. All selective episodes (or, perhaps, almost all) can be interpreted in terms of genic selection. That is an important fact about natural selection. It is not, as Sober appears to think (*ibid.*, p. 151), something that is quasi-analytic on our view. That genic selection could be found in hard cases—those addressed by Richard Dawkins, John Maynard Smith, and (to a far smaller extent) ourselves—is surely on any view a major discovery.

Furthermore, the question whether there are real examples of processes that can be modeled as group selection can be asked and answered entirely within the genic point of view. Whereas Sober prefers to describe the possibility in terms of group selection, genic selectionists can couch it in terms of the existence of special genetic environments and/or populations. The empirical issue about the existence of scenarios of the type Sober imagines does not vanish within the genic image, but is recast in different terms.[2]

We have pointed out that genic selectionists have to be careful in characterizing populations and environments—for some genic accounts see selection as context-dependent—but there is no bar to taking the gene's-eye-view.[3] Nor is there any point in insisting that selection is really "context-independent higher-level selection" rather than "context-dependent lower-level selection." To justify that insistence, one requires a clear criterion for distinguishing between accounts of selection processes that are context-dependent and those which are not—thus identifying those causal profundities which shallow pluralists miss.

Sober used to have a crisp formulation of the causal commitments of selection models. The account, given in *The Nature of Selection*, proposed that the units of selection are those things which bear the properties for which selection occurs, and explicated selection for a property by demanding that the presence of the property raise fitness in at least one relevant context and not lower it in any. On the basis of the account, genic selection was supposed to be demonstrably inadequate. We demur. We have argued that Sober's explication of *selection for a property* will not sanction everyday (paradigm!) cases of natural selection, since there are typically microenviron-

[2] See C. Kenneth Waters, *Models of Natural Selection: From Darwin to Dawkins* (doctoral dissertation, Indiana University, 1986); and "Tempered Realism about the Forces of Selection" (forthcoming).

[3] See Kim Sterelny and Philip Kitcher, "The Return of the Gene," this JOURNAL, LXXXV, 7 (July 1988): 339–361; and Waters, *op. cit.*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ments encountered by members of the population in which the probabilistic inequalities are reversed.[4] We further pointed out that elaboration of the notion of an allelic environment would enable claims about selection for genic properties to conform to Sober's principle.

Sober sees our position as impoverished; not so. The fact that the genic perspective gives us a general picture of selection processes is an important idea. The extent to which individual- and group-level descriptions are available remain open empirical questions. Sometimes, however, scientists, as well as philosophers, debate pseudo problems. Once the possibility of many, equally adequate, representations of evolutionary processes has been recognized, philosophers and biologists can turn their attention to more serious projects than that of quibbling about the real unit of selection. The riches accruing to those who insist on the right to quibble are purely illusory.

PHILIP KITCHER
University of California/San Diego

KIM STERELNY
Victoria University of Wellington

C. KENNETH WATERS
Rice University

BOOK REVIEWS

Descartes. MARJORIE GRENE. Minneapolis: University of Minnesota Press, 1985. 225 p. Cloth $35.00, paper $14.95.

Marjorie Grene displays in *Descartes* three distinctive trademarks of her distinguished career in philosophy. First, and above all, she shows her great capacity to engage in a philosophical dialogue with a past thinker which is both sympathetic and combative. In order to achieve such a spirited dialogue, she mobilizes the multiple skills of a historian, metaphysician, and philosopher of science. This second trademark of her work has as its purpose not only the fashioning of a full treatment of her subject matter but also the development of the philosophy of tacit knowledge, for which, following Michael Polanyi, she has been a principal spokesperson. Readers of *Descartes* will be pleased to discover that, in keeping with her third trademark as an

[4] See Sterelny and Kitcher, *op. cit.*, and Waters, *op. cit.*

0022-362X/90/8703/161–5
© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

exponent of tacit knowledge, she has found new avenues to explore in the psychological theories of J. J. and E. J. Gibson.

Grene, as historian and metaphysician, brings together in this book the consideration of several classic French interpretations of Descartes, as well as a review of an impressive variety of recent Anglo-American studies that have treated Descartes and his philosophical contemporaries as serious subjects of historical inquiry. Pivotal to her assessment of Cartesian philosophy is the work of Martial Gueroult,[1] who in his *Descartes selon l'ordre des raisons* sought to give a rigorous explanation of the argumentative aims of the *Meditations* in terms of its structure. Grene endorses Gueroult's interpretation of this structure, characterized by him as a diptych in which the metaphysical arguments comprising the first panel require the Cartesian meditator to conceive of himself through the understanding as constituting mind alone, but in which the metaphysical arguments of the second panel prescribe a return to the cultivation of knowledge of the senses. This return from the occupations of the understanding to those of the imagination and the senses is, for Gueroult, induced by the meditator's recognition that, despite the real distinction between mind and body, he has been created as a composite of these two substances by a veracious and benevolent God (*ibid.*, vol. 2, pp. 130, 217–9). Gueroult argues that Descartes's exploration of the incoherent consequences of the mind-body union should be viewed as the deliberate identification by a rationalist philosopher of the "irrational elements that we believe we are capable of discovering in the work of God" (*ibid,* p. 237). His Descartes applies the principle of the best to justify a divinely ordered world in which human beings must cope with their composite natures and the obscurity and confusion of their sense ideas. Grene's reading of the *Meditations* parallels Gueroult's in its emphasis on the significance of Descartes's account of sense perception and the incoherence of his notion of a mind-body union. She interprets the text not as a Cartesian theodicy, however, as Gueroult has done, but rather in terms of what she considers was the deleterious influence of the way of ideas on the subsequent development of modern philosophy. Grene's Descartes commits the crucial error of equating his bodily encounters with the world with his possession of ideas of sense, and as a result he attributes to such encounters an obscurity and confusion largely of his own making. This error and Descartes's

[1] *Descartes' Philosophy Interpreted According to the Order of Reasons,* 2 vols. Roger Ariew, trans. (Minneapolis: Minnesota UP, 1984–85).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

concomitant difficulties in specifying mind-body relations are trenchantly criticized by Grene, who describes them as symptomatic of the philosophical shortcomings of the Cartesian way of ideas (14–21, 30–52).

In her own historical treatment of the reception of the *Meditations* by Descartes's contemporaries, Grene deftly draws up a balance sheet of what she believes was gained and lost by philosophers who adopted various Cartesian innovations in epistemology and metaphysics. She also notices which of these innovations seem to have been systematically misunderstood by Descartes's seventeenth-century rivals. It is in this portion of her book that Grene generously showcases the work of a number of recent historians of philosophy whose research she incorporates into her own discussion of the responses to Descartes of Mersenne, Gassendi, Arnauld, the Jesuits, and other Aristotelians. When, for example, the Aristotelian critic Caterus complains that whatever Descartes refers to as an "idea" is unintelligible, or when Gassendi objects to what he views as Descartes's surreptitious reliance on a misconstrual of the scholastic doctrine of natural light, Grene underscores the opacity of these Cartesian innovations to both advocates and opponents alike of scholastic orthodoxy (124, 144–8). She also explicates their strikingly similar reservations concerning his rejection of the primacy of sense experience in the acquisition of knowledge and the attenuated role that he allots to judgments in justifying the truth of his beliefs. Grene's balance sheet weighs, on the negative side, these two features of his work against, on the positive side, his arguments demonstrating the ineffectiveness of the scholastics' substantial forms doctrine in physics (53–4, 95–9, 195–6). But even here she adds that, in the biological sciences, an Aristotelian concept of substance might have led to better theories if Descartes had never suggested a two-substance metaphysics to fellow scientists.

Grene's sympathies clearly lie with the seventeenth-century as well as twentieth-century opponents of Descartes's epistemology and metaphysics, although she distinguishes her anti-Cartesian stance from, among others, that of Richard Rorty, and although she certainly acknowledges the work of contemporary analytic philosophers such as Harry Frankfurt, E. M. Curley, and Margaret Wilson, who are more sanguine about Descartes's achievements than she is (4, 212). She occasionally paints with a very broad brush, however, and some readers will question, for instance, whether her reasons for holding Descartes responsible for the excesses of sense-data theorists of perception are in fact good reasons. These readers will further

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

note the absence in her book of any extended discussion of two philosophers in the analytic tradition who made major contributions to the dispelling of just those later manifestations of Cartesian beliefs which she judges to have been most harmful. Wilfrid Sellars'[2] powerful attack on sense-data theorists and their credulousness concerning the myth of the given surely deserves mention in Grene's account of the fortunes of Cartesianism in our time, as does Gilbert Ryle's[3] exorcizing of the mental ghost in Descartes's man machine. Grene is perhaps most puzzling when, in commenting on the present state of affairs in philosophy, she continues to see the ill effects of Descartes's influence *even after* so many developments that signal the waning of that influence.

The depth of Grene's alienation from what she regards as the persisting adherence to Cartesian modes of philosophizing in contemporary philosophy emerges in her final essay, in which she contrasts Descartes's theory of sense perception with that advanced by the psychologist J. J. Gibson in *The Ecological Approach to Visual Perception.*[4] She attempts to show, in her discussion of Gibson's theories of affordances and invariants, what an account of perception must do in order to rid itself of all Cartesian presuppositions. The theory of affordances, she quotes Gibson as saying, interprets "sense-mediated information . . . as 'properties of things taken with reference to an observer, but not properties of the experiences of the observer' " (209). Gibson's invariants are then illustrated by recalling how "a child understands what an object is before it grasps reflectively the concept of something extended" (209). Finally, Grene links Gibsonian psychology to the philosophy of tacit knowledge: "In the *perception* of identifiable objects, and the events through which they persist or alter, we find in fact the tacit foundation of *all* knowledge . . ." (209–210).

Her ultimate disagreement with Descartes thus turns on the incompatibility of tacit knowledge not just with his account of sense perception, but also with his insistence on the explicit, conceptual character of human knowledge. Here Grene's interpretive sense rings true. What was genuinely remarkable about Descartes was his ability to preserve the strongest commitment to the conceptual character of human knowledge even as he redefined that character in terms of ideas rather than judgments or propositions. For while he

[2] "Empiricism and the Philosophy of Mind," in his *Science, Perception and Reality* (New York: Routledge & Kegan Paul, 1963), pp. 127–96.
[3] *The Concept of Mind* (London: Hutchinson, 1949), pp. 11–61.
[4] New York: Houghton Mifflin, 1979.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tried to privatize our thoughts and to give to each mind a privileged access to its own ideas, he never encouraged anyone to believe that the foundations of knowledge might be nonconceptual, tacit, and exemplified best by the prelinguistic learning of infants. His most radical notion, the epistemological transparency of the mind to itself, still presupposed fluency in *some* language. In her feisty, engaging book, Grene asks us to consider whether we dare to think that Descartes was wrong about this, too. Whether or not we agree with her specific proposals regarding Gibsonian psychology, she warns that in the balance of our answer to her larger question hangs the future of philosophy.

LYNN S. JOY

University of Notre Dame

NOTES AND NEWS

The Centre for Religious Studies at the University of Toronto is pleased to announce a new journal, *Method and Theory in the Study of Religion*, which provides a forum for scholars in a number of disciplines to discuss and debate issues of method and theory in the academic study of religion. The journal is published semi-annually at the Centre, in association with the North American Association for the Study of Religion. The first issue appeared in Spring 1989. The journal accepts articles, critical notes, and book reviews, with an effort to reserve a majority of space for younger scholars. Further information may be obtained by writing Ann Baranowski, *Method and Theory in the Study of Religion*, Centre for Religious Studies, 130 St. George St., Univ. of Toronto, Toronto, Ontario M5S 1A5, Canada.

Essays are being sought for a volume exploring connections of gender to traditional problematics of knowledge, including truth, justification, belief, agency, and power. Also welcome are critiques of traditional epistemologies as well as new feminist epistemologies. Further information may be obtained by writing Linda Alcoff, Phil. Dept., Syracuse Univ., Syracuse, NY 13244, or Elizabeth Potter, Phil. Dept., Hamilton College, Clinton, NY 13323.

The Society for the Philosophy of Sex and Love will be meeting with the Eastern Division of the American Philosophical Association in December 1990 (Boston) and with the Central Division of the APA in April 1991

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(Chicago). Philosophical papers on sexuality or love, twelve pages in length (excluding footnotes), and in duplicate, may be sent to the address below by April 1 (for the Eastern meeting) or by October 1 (for the Central meeting): Alan Soble, Phil. Dept., Univ. of New Orleans, New Orleans, LA 70148.

Teaching Philosophy is offering a $250 Prize for the best paper submitted for a special issue of the journal devoted to teaching modern philosophy. The deadline has been extended to June 30. Three copies of submitted papers may be sent to Arnold Wilson, Editor, *Teaching Philosophy,* Univ. of Cincinnati, Cincinnati, Ohio, 45221-0206. Also, the Central Conference on Teaching Philosophy meets annually with the APA Central Division. For a program on Cultural Pluralism and Diversity in the Curriculum, the Conference is soliciting papers on the role of philosophy in this growing interest, or the significance of it for philosophy teaching and curricula. Selected papers will be presented at the annual meeting, and the best will win a $250 Prize offered by *Teaching Philosophy* and will be published therein. Papers and inquiries may be addressed to Martin Benjamin, Phil. Dept., Michigan State Univ., East Lansing, MI 48824.

The Institute for Vico Studies announces the sixth annual issue of *New Vico Studies,* edited by Giorgio Tagliacozzo and Donald Phillip Verene. *New Vico Studies* publishes articles, reviews, and notes that reflect the current state of the study of the thought of Giambattista Vico, including ideas by others that are Vichian in nature. Subscription orders may be placed through Humanities Press International, Atlantic Highlands, NJ 07716.

The Department of Philosophy at Carnegie Mellon University is pleased to announce the appointments of Cristina Bicchieri as Associate Professor, David Carrier and Wilfried Sieg as Professors, Teddy Seidenfeld as Head, and Clark Glymour as Alumni Chair in Philosophy; and the return of Preston Covey as Associate Professor and Director of the Center for the Advancement of Applied Ethics.

The Department of Philosophy at Indiana University (Bloomington) is pleased to announce the appointments of Jon Barwise as College Professor of Philosophy, Mathematics, and Logic, Anil Gupta as Professor, Alice ter Meulen as Associate Professor of Philosophy and of Linguistics, and Frederick Beiser and Timothy van Gelder as Assistant Professors. We wish also to announce the availability of a new fellowship for graduate study in logic. Applicants may apply to any appropriate department—for example, Philosophy, Mathematics, History and Philosophy of Science, Computer Science, Linguistics, or Psychology. Requests for more detailed information about the fellowship may be obtained by writing to Linda Harl, Phil. Dept., Sycamore Hall 026, Indiana Univ., Bloomington, IN 47405.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The Department of Linguistics and Philosophy at the Massachusetts Institute of Technology is pleased to announce a conference on the history and philosophy of science in honor of Thomas Kuhn. The sessions will take place at MIT on May 18–19. Participants will include Jed Buchwald, Nancy Cartwright, John Earman, Arthur Fine, Michael Friedman, Ian Hacking, John Heilbron, C. G. Hempel, Thomas Kuhn, Ernan McMullin, Noel Swerdlow, and Norton Wise. More information may be obtained from Paul Horwich or Judith Thomson, Dept. of Linguistics and Phil., 20D-213, MIT, Cambridge, MA 02139.

The Department of Philosophy at the University of California/Riverside is pleased to announce the appointments of John Martin Fischer as Associate Professor, Pierre Keller and Allison Shalinsky as Assistant Professors, and Howard K. Wettstein as Professor of Philosophy.

The Department of Philosophy at the University of Colorado/Boulder is pleased to announce the appointments of Alison Jaggar as Professor of Philosophy and Women Studies, and Robert Hanna and Michael Otsuka as Assistant Professors.

The Department of Philosophy at the University of Massachusetts/Amherst is pleased to announce the appointment of Lynne R. Baker as Professor of Philosophy.

The University of Pittsburgh is pleased to announce that Adolf Grünbaum, the Andrew Mellon Professor of Philosophy, Research Professor of Psychiatry, and Chairman of the Center for Philosophy of Science, has been awarded the 1989 Fregene Prize for Science, in recognition of his distinguished career as a philosopher of science. The Prize was initiated five years ago in Rome, and is awarded annually to recognize outstanding career achievements in science.

The Department of Philosophy at Virginia Polytechnic Institute & State University will hold its annual conference March 30–April 1. The topic for this year's conference is: "The Role of Experimentation in Scientific Change." The participants include: Peter Achinstein, Roger Ariew, Peter Barker, Robert Brandon, Jed Buchwald, Richard Burian, Peter Dear, Gary Downey, Mordechai Feingold, Steve Fuller, Peter Galison, Daniel Garber, Marjorie Grene, Elsworth Furhman, Clark Glymour, Donald Howard, Bruno Latour, Ronald Laymon, Deborah Mayo, Joseph Pitt, and Simon Schaffer. Further information may be obtained by writing Joseph Pitt, 309 Patton Hall, Virginia Polytechnic Institute & State Univ., Blacksburg, VA 24061-0126.

The Department of Philosophy and Religious Studies of Youngstown State University, is pleased to announce the appointment of Stanley M. Browne as Assistant Professor of Philosophy.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the philosophical forum
A QUARTERLY

VOLUME XXI:1-2 $8.50
Fall-Winter 1989-90

HERMENEUTICS IN ETHICS AND SOCIAL THEORY
A SPECIAL DOUBLE ISSUE

Jürgen Habermas
Michael Walzer
Agnes Heller
Thomas McCarthy
Seyla Benhabib
Albrecht Wellmer
and others

Editor: Marx Wartofsky
Guest Co-Editor:
Michael Kelly

SUBSCRIPTIONS
$15 per year for individuals
$45 per year for institutions
+ $4 for foreign subscriptions

Other special issues:
Latin American Philosophy
Fichte
Apartheid

the philosophical forum
P.O. Box 239
Baruch College of CUNY
17 Lexington Avenue
New York, NY 10010

Cognitive Economy
The Economic Dimension of the Theory of Knowledge
Nicholas Rescher

In outlining his theory of the cost-effective aspects of research, Rescher shows that the rational pursuit of maximum benefits needs to be viewed as a central concern with respect to the use of intellectual resources.

"Rescher at his best. The insights are deep, the arguments are strong and full of imagination, the applications of the model are precise and apt. . . . The book should be required reading for all those who serve on grants committees."—Robert Butts, The University of Western Ontario

$29.95/ISBN 0-8229-3617-8

Pittsburgh
University of Pittsburgh Press
Pittsburgh, PA 15260

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVII, NO. 4, APRIL 1990

THE REWARD EVENT AND MOTIVATION*

IN philosophy, the textbook case for the discussion of human motivation is the examination (and almost always, the refutation) of psychological egoism. The arguments have become part of the folklore of our tribe, from their inclusion in countless introductory texts.

The arguments are too familiar to need repetition. It is often concluded that psychological egoism is either confused, or involves a misuse of language, or collapses into tautology or infinite regress, or, if barely empirical, is just obviously wrong.

I am not interested in egoism as such, nor in the strictly logical or linguistic arguments against psychological egoism in its more confused forms. What I am interested in are certain arguments used in the refutations which seem on the face of it to be making significant empirical claims about motivation. Here are two representative examples.

The first is from James Rachels'[1] discussion of psychological egoism:

> There is a general lesson to be learned here, having to do with the nature of desire and its objects. If we have a positive attitude toward the attainment of some goal, then we may derive satisfaction from attaining it. But the *object* of our attitute is *the attaining of that goal;* and we must want to attain the goal *before* we can find satisfaction in it. We do not first desire some sort of "pleasurable consciousness" and then try to figure out how to achieve it. Rather, we desire all sort of different things . . . and *because* [my emphasis] we desire those things we derive satisfaction from getting them (*ibid.*, pp. 59–60).

* Thanks to Norton Nelkin, Jim Stone, Ed Johnson, and all the Friday regulars for very helpful comments.

[1] *The Elements of Moral Philosophy* (New York: Random, 1986).

0022-362X/90/8704/169–186

© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The second is from Joel Feinberg's[2] discussion of psychological egoism, starting with his discussion of the "Lincoln and the baby pigs" case:

> The very fact that he [Lincoln] did find satisfaction as a result of helping the pigs presupposes that he had a *preexisting* [my emphasis] desire for something other than his own happiness . . . The *object* of Lincoln's desire was not pleasure; rather pleasure is the *consequence* of his preexisting desire for something else . . . But if pleasure and happiness *presuppose* [my emphasis] desires for something other than pleasure and happiness, then the existence of pleasure and happiness in the experience of some people proves that those people have strong desires for something other than their own happiness—egoistic hedonism to the contrary (*ibid.*, p. 532).

It should be emphasized that neither Rachels nor Feinberg is at all confused about the difference between satisfaction in the sense of obtaining what one desires, in which sense it is trivially true that when we get what we want we are satisfied, and a very different sense, intended to focus on an aspect of experience that is only contingently connected with the getting of what we want. As Feinberg puts it: "Fulfillment of desire (simply getting what one was after) is no guarantee of satisfaction (pleasant feelings of gratification in the mind of the agent)" (*ibid.*, p. 531).

Although I do not want to be committed to that particular characterization of the internal psychological event on which both Rachels and Feinberg focus, it is clear that they both have in mind some such event. For reasons that will become clearer as we proceed, let us call it the *reward event*.

Both Rachels and Feinberg seem to be making the following claims:

(1) In most (though not all) cases of human motivation, the ultimate object of the motivation or desire, what the agent wants for its own sake, is some external object or state of affairs, not some internal reward event.
(2) There is a variety of such ultimate objects.
(3) In many cases of such outwardly focused motivation, if the object is realized and the agent knows that it is, this may, as a contingent fact, cause some reward event in the agent, but in such cases the reward event will occur only if there is already a preexisting desire for some object other than that internal reward event; the desire is not causally dependent on the reward event.

[2] "Psychological Egoism," in Feinberg, ed., *Reason and Resonsibility* (Encinco, CA: Dickenson, 1978, 4th ed.).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Let us call this the *by-product theory*. According to this theory, human motivation is (at least in many cases):

(1) outwardly focused,
(2) pluralistic,
(3) temporally and causally prior to any associated internal reward event.

It seems further to be assumed that, once one has cleared up the logical and linguistic confusions in egoism, it will be obvious that this outward, pluralistic view of motivation is correct.

But if a theory of human motivation is empirical, is really making claims about how we "work," can it be that easy? If one is sympathetic to the naturalistic enterprise, then one will expect that an understanding of motivation will need to be as responsive to scientific developments as must our understanding of perception.

Many have recognized this, though not often in the context of a discussion of egoism. One notable exception is Michael Slote,[3] in an article reprinted in Feinberg's own introductory anthology (through the 4th ed.).

Slote argues that "psychological egoism may well have a basis in the empirical facts of human psychology" (*ibid.*, p. 540). The core of his case goes as follows:

> Certain contemporary learning theorists, e.g., Hull and Skinner, have put forward behavioristic theories of the origin and functioning of human motives which posit a certain number of basically "selfish," unlearned primary drives or motives (like hunger, thirst, sleep, elimination, and sex), and explain all other, higher-order, drives or motives as derived genetically from the primary ones via certain "laws of reinforcement," and further deny the "functional autonomy" of the higher-order desires or motives . . . We have reason to believe that a given higher-order drive or motive is *not* functionally autonomous, i.e., is functionally dependent, if when we cut off all reinforcement of it by primary rewards (rewards of primary drives) and there are, in addition, both a sufficient number of "extinction trials" (occurrences of acts done from that higher-order motive which are not associated even indirectly, i.e., through other higher-order motives, with primary rewards) and a complete absence, during these extended trials, of primary rewards for any similar higher-order motives (to eliminate the possibility of generalization of primary rewards from motives other than that being extinguished), the higher-order drive or motive actually does extinguish, i.e.,

[3] "An Empirical Basis for Psychological Egoism," this JOURNAL, LXI, 18 (October 1, 1964): 530–7; repr. in Feinberg, ed., 540–4. References will be to the page numbers in the Feinberg volume.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the person whose higher-order motive is being extinguished eventually, even if perhaps only very gradually, ceases to act from that higher order motive (*ibid.*, pp. 540–1).

Admittedly, when one considers all the conditions placed on the extinction trials required by the theory, and tries to imagine their being fulfilled for humans, this seems rather tenuous; but we need only admit that such a theory could be true. Nor am I much concerned with whether we would want to call such a theory egoistic. Comparing this behavioristic learning theory of motivation (BLTM) to the by-product theory is a preliminary to my main project.

(1) Since the background theory for BLTM is behaviorism, which countenances no internal psychological events,[4] the focus of motivation will be outward. The ultimate objects of motivation will be, however, much more severely restricted to such things as food, water, sex object, or perhaps eating food, drinking water, copulating. Only unlearned, and in that sense primitive, objects will be on the list.

(2) The theory is thus minimally pluralistic, unlike the wide-ranging pluralism of the by-product theory.

(3) Since the primary drives or motives are unlearned, that is, genetically wired in, the theory seems to assume that these are temporally and causally prior to any primary rewards; some things (food, water, etc.) are rewards because we have primary drives for them.

A few comments are in order. First, for BLTM to be at all plausible, it must admit that at the level of everyday conscious experience we are indeed focused on the great variety of objects that the by-product theory emphasizes; they are what we take to be our goals. But that is no problem, since the issue has always been one of ultimate goals, that is, those things we pursue for their own sake and not merely as means to some other end. And if one is not committed to the view that our everyday understanding of our aims and motives is necessarily accurate, we should perhaps expect to be surprised by what a theory of motivation tells us. So what BLTM is saying is that, while we are indeed *focused* on a great variety of goals, all of these

[4] For historical accuracy, it should be noted that Clark Hull developed a "*drive* theory of reinforcement and learning in which bodily need led to drive, which then led to new learned behavior to correct the need and reduce the drive, thus reinforcing the learning" [as discussed in James R. Stellar and Eliot Stellar, *The Neurobiology of Motivation and Reward* (New York: Springer, 1985), p. 23]. This does seem to give a primary role to internal conditions in the organism, but these are internal physiological conditions, and, to the degree that the psychological terms, e.g., 'drive' and 'reinforcement', are defined operationally in terms of patterns of behavior, it is still in the spirit of behaviorism.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

are *anchored* in, derive their motivational force from, a few unlearned primary goals. And, of course, any defender of the by-product theory would have to admit that most of the goals of adults, the goals which are at issue, are learned. We are not born desiring to listen to Mozart or to corner the market in silver. That is why the claim that such learned goals never become functionally autonomous is the empirical heart of BLTM; it is a theory not merely about how we first came to learn new goals, but about what sustains them as goals.

Second, despite their differences, BLTM and the by-product theory agree that desire or motivation is causally primary. In all (BLTM) or most (by-product theory) cases, a reward is a reward, is that which contingently satisfies us, because we desire it, we do not desire it because we have independently found it to be satisfying.

Third, both theories radically downplay the role of internal psychological reward events. BLTM, of course, officially denies there are any such events, or at least any we can or should include in a scientific psychology. The by-product theory, while allowing that there are such events (e.g., some kind of felt satisfaction) and that they do play some role in our psychology, must insist that, with respect to much of adult motivation, such events are not causally primary but are contingent by-products of our autonomous desires for all sorts of things other than such internal events. This is, after all, the heart of the case against psychological egoism.

I have stressed similarities between the by-product theory and BLTM, because the chief aim of this paper, given Slote's point that it is an empirical question which theory of motivation is correct, is to sketch an alternative theory that (1) would be anchored in internal reward events, (2) would be, in a sense, strongly monistic, and (3) would reverse the causal order, insisting that we intrinsically desire these reward events because we find them to be intrinsically satisfying—we do not find them to be intrinsically satisfying because we intrinsically desire them.

The theory will be based on some empirical straws in the wind, plus my own sense of where psychological explanation may be heading. Admittedly the superstructure I erect is highly speculative, but all I wish to claim is that it has some empirical support, that it might be true, that we are in no position at present to rule it out, and that, though it may not itself be a form of egoism (that will remain to be discussed), it does rule out many of the standard arguments against egoism as an empirical theory of human motivation. In addition, and of more general significance, the theory provides a most suggestive framework for thinking about motivation or desire, the objects of

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

motivation and desire, intrinsic reward or satisfaction, and the respective roles of each of these in the lives of many (though perhaps not all) natural creatures.

So let us sketch the theory, and then consider its implications. I shall call it the *reward event theory* (RET).

I

There are two foundations for the theory, one rigorously experimental and dealing with the rat, the other impressionistic and anecdotal and dealing with humans. In between is a fair amount of neurobiological and evolutionary speculation.

The first foundation is the enormous literature concerning the reward effect, in rats, of electrical stimulation of the brain (ESB). Since this is to be an empirical foundation, I shall quote rather extensively from a recent overview of the literature—Stellar and Stellar, *The Neurobiology of Motivation and Reward* (NMR)—and its significance for wider psychological issues.

The classic experiments were performed by Olds and Milner in 1954, "in which they demonstrated that electrical stimulation of parts of the limbic system (septal area and diencephalic regions) of the rat had a powerful reenforcing effect in that it shaped the rat's lever-pressing behavior and made it respond at an extremely high rate" (NMR, 18). Later experiments demonstrated that ESB "was a very strong motivated behavior, for rats took very strong foot shock in order to stimulate their own brains (Olds, 1958c)" (NMR, 31–32).[5] Indeed, the stimulation has been shown "to have a suppressive effect on several pain-motivated withdrawal behaviors . . . During lateral hypothalamic stimulation, rats fail to respond to strong ammonia-soaked swabs, which cause trigeminal pain in the olfactory membranes and they also show attenuated startle to intense auditory stimuli (Stellar, 1976)" (NMR, 203). Other experiments demonstrated "the strong preference for ESB over food, even to the point of starvation (Routtenberg & Lindy, 1965) . . ." (NMR, 196). We might also note that early on "Heath (1964b) and his colleagues . . . extended these finding to humans and were able to elicit reports of hedonic experience upon brain stimulation in the same regions" (NMR, 22). Although the literature speaks varyingly of pleasure, reenforcement, and reward, it should not beg too many questions right at the start to speak generically of whatever is going

[5] It will be clear, from many of the quotations I shall be using, that these experiments raise significant ethical issues concerning our treatment of nonhuman animals. I have read the literature with an uneasy combination of intellectual fascination and ethical concern. But (is this the intellectual's easy out?) that is an issue for another time.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

on in the rat's brain as the reward event. Thus the name of our theory.

The second foundation, more tenuous and anecdotal, consists of a number of introspective reports by those who have become suddenly and overwhelmingly addicted to the smoking of "crack," a particularly potent form of cocaine. The individuals report that the initial experience is overwhelmingly pleasurable (rewarding) and thereafter dominates them psychologically. They think about it constantly and want nothing so much as to try to repeat the experience. And it is not merely those who, as we say, did not have much else going for them psychologically, who find themselves gripped by this dominant preference. Those with a wide variety of interests, successful careers, any number of those outwardly directed interests emphasized by the by-product theory, report a similar psychological dominance.[6] In addition, the neurologists tell us that the swiftness of the addiction is due in part to the fact that, in smoking crack, the cocaine goes right to the brain. Finally, in order to have a few more empirical lines out, we may note that there is some evidence for a connection between the ESB reward event and the neurotransmitter dopamine, and cocaine is a dopamine-stimulating drug. Rats will self-administer cocaine (much like ESB), but the self-administration is blocked by dopamine blockers. There is also evidence that opiate reward works through a dopamine mechanism (cf. NWR, 184–6, 193).[7]

Those are the foundations. Now for the superstructure. Remember, it is speculative, though I shall at certain points indicate where there is some empirical support. I shall also try to make clear where there are great gaps in the evidence, and where existing evidence could well be read in ways antithetical to the theory. A theory of motivation, if properly empirical, should be open to refutation.

Let us suppose that it turns out that the brain events in the bar-

[6] For a recent update on research concerning drug addiction, in both animals and humans, see E. J. Engel and L. Oreland, eds., *Brain Reward Systems and Abuse* (New York: Raven, 1987).

[7] Cf. also, R. A. Wise and M. A. Bozarth, "Action of Drugs of Abuse on Brain Reward Systems: An Update with Specific Attention to Opiates," *Pharmacology Biochemistry & Behavior*, XVII (1982): 239–243: "While additional work is needed and is in progress, the current data thus suggest that opiates are rewarding because of a local activation of dopaminergic cells in the central tegmental area; these cells are also presumed to play a critical role in brain stimulation reward, psychomotor stimulant reward and *food and water reward*" (241, my emphasis).

There is an enormous experimental literature on drugs and addiction, much of it still dealing with the rat, but much also focused on humans, so this foundation for RET is "more tenuous and anecdotal" only in involving introspective reports and the enormous complexities of human motivation, making theoretical interpretation much more difficult.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

pressing rat and those in the crack-smoking human are sufficiently similar neurobiologically to justify speaking, at the type level, of *the* reward event. Next let us suppose that as there is further investigation of what goes on in the brain when a creature finds food, or water, or sex, or listening to Mozart rewarding, it turns out that in all cases there is an instance of the reward event, and that, in the absence of the reward event, these things are not rewarding. There are experiments that suggest that ESB reward has some of the properties of such "natural" rewards as food or water (cf. NMR, 78, 200).[8] And Stellar and Stellar rather boldly hypothesize that "all motivated behaviors and associated hedonic processes share similar neurological mechanisms in the limbic systems and that all operate according to the same basic physiological and behavioral principles" (NMR, 50). But they are also quick to caution that there is as yet little conclusive evidence linking ESB reward and other forms of motivation, even those much-studied ones of hunger and thirst. "The problem of relating ESB reward at any site to natural rewards is even more difficult" (NMR, 154).

Next, let us suppose that, in most instances, learning takes place only if the reward event is triggered in the brain.[9] And finally, let us assume that learned motivation remains functionally dependent (in Slote's sense) on connecting with the reward event at appropriate intervals. Thus, "higher-level" learned motivations would have their causal origin in, and would be causally sustained by, the reward event.

The reward-event theory anchors all *positive* motivation in the reward event. The qualification is important. Up until now we have been speaking as if all motivation were aimed at gaining some goal. But of course creatures can be, and undoubtedly are, motivated to avoid some things. Let us bracket these phenomena, or presume, in the spirit of RET, that all such aversion-based learning is also internally anchored, in some aversion event (or events). This preserves the theory as strongly internal; that is, a creature is motivated to act in a certain way if and only if there is some appropriately associated

[8] Cf. also the Wise and Bozarth quotation in fn. 7.

[9] There is already considerable evidence that this is not true of all cases of learning, that in many if not most creatures there is considerable "latent learning," associations learned in the absence of any reward. Of course, the experiments were considering the "traditional" rewards of food, water, etc.; it is much more difficult to control for the absence of a kind of brain event. But there may be deep differences between motivated and unmotivated learning, and since my concern is with motivation, we can let that pass. Cf. NMR, 18: "Thus, even if motivation and reward are not necessary conditions for all classical conditioning, they often accompany it. In operant conditioning, of course, motivation and reward are clearly two conditions that must prevail if the law of effect is to apply."

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

internal event, rewarding or aversive.[10] Hereafter, we shall be dealing only with positive motivation and the reward event.

This is the outline of the theory. I shall now try to characterize it further, draw out some of its implications, and, along the way, consider objections either to the possibility of such a theory turning out to be true, or to what its implications would be if it were true. This should put some philosophical flesh on these meager empirical bones.

II

I shall proceed by discussing RET with respect to each of the three issues in terms of which we characterized the by-product theory and BLTM—(1) internalism versus externalism of the ultimate objects of motivation, (2) monism versus pluralism of those objects, and (3) the direction of causation between motivation and reward, or desire and satisfaction.

(1) *internal or external.* One of the least plausible aspects of psychological egoism was its insistence that the only ultimate objects of motivation, what we want for its own sake, are certain of our own psychological states. To some degree we dealt with this in our discussion of BLTM by distinguishing between the *focus* of motivation and the *anchor* of motivation, but the challenge is still there, and could be phrased as follows. Even if something like RET is true, why does that not merely tell us more about the mechanism of motivation, about why we have the many different objects of motivation we do have? Why should that mechanism itself count as the only ultimate object? Several comments are in order.

First, we should remember the peculiar dominance of the ESB experience in the rat and the crack experience in the human. It is as if these experiences, these reward events, cut loose from any distracting association with external objects or events, were for the first time revealed in their full motivational strength. Thus, the operation of the reward event in more mixed phenomena may be buffered in such a way that focus shifts; but the reward event would still be the aspect of these more complex experiences which is what we are motivated to obtain. There is some empirical support for this speculation. "[Head and Holmes (1911)] reported clinical cases with thalamic lesions in which stimuli were either excruciatingly painful or exquisitely pleasant, and these included as diverse sources as light tactile stimulation (pain) and music (pleasure). They speculated that

[10] Cf. NMR, 18: "Olds and others then went on to show that stimulation in some loci in the brain was aversive, in that it diminished and eliminated operant behavior (Olds & Olds, 1963; Valenstein & Valenstein, 1964; Hoebel, 1976). *It was as though the stimulating electrode had artificially activated parts of the brain that natural positive and negative reinforcers activate*" (my emphasis).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the thalamus is normally *inhibited by the cortex in the expression of hedonic experience*" (NMR, 20, my emphasis).

Second, there are good evolutionary reasons why our focus should be outward. A creature could neither survive nor evolve if it had a dominant preference for some internal psychological state that it could somehow produce directly, instead of through the obtaining of food, water, sex, etc. The ability to "mainline" reward events is something of an artifact, which may also explain why creatures have not evolved defenses against its dominance when it is available.

Third, there are advantages to having motivation depend upon reward events that are only contingently connected with external events and objects. This could well be the motivational key to flexibility of response and to the learning of new goals. There may be whole organisms, or kinds of behavior within organisms, where response is directly to some perceived stimulus, without any connection through reward events, and in a certain sense such behavior would have intrinsically outward objects. But the best examples of that would not be high-level human motivations but reflex responses, such as the toad's zapping small moving objects ("bugs") with its tongue when they are detected visually. This suggests something of a challenge: What would our neurology need to be like for purely cognitive states, e.g., belief that certain actions could help others, to be intrinsically motivating, instead of depending on some connection with some affective states? Is there any evidence to support such "Kantian" motivation?

Fourth, the contingency of connection between the ultimate goal of the reward event and sundry external objects is also suggested by several human phenomena. One of the most familiar is what could be called the *Wordsworth effect*—the gradual, or sudden, loss of the sense of intrinsic delight in scenes and activities that formerly were strongly motivating, say, the natural landscapes of childhood. One is tempted to postulate a neurological answer to the question posed in the Thurber cartoon: "But who made the magic go out of our marriage?" At a greater extreme, there are the cases of extreme depression, often part of the manic-depressive or bi-polar syndrome. In a recent television program on this disorder, one woman reported that while in the depression phase she could not even get out of bed to feed her five children. Might a massive failure of connections with reward events account for such a massive failure of motivation?[11]

[11] Some current theories of depression can be read this way. A recent article in the Health Letter of the New York University Medical Center identified the characteristic symptoms of major depression as insomnia; poor appetite, often accompanied by weight loss; loss of interest in hobbies, other people, and outside events; guilt feelings, and self-destructive thoughts. A clinical instructor of psychiatry at the Medical Center, Dr. Antoinette A. Wysznski, is quoted as saying: "Especially promi-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Finally, there are the related phenomena explicitly noted by Feinberg:

> Fulfillment of desire (simply getting what one was after) is no guarantee of satisfaction (pleasant feelings of gratification in the mind of the agent). Sometimes when we get what we want we *also* get, as a kind of extra dividend, a warm, glowing feeling of contentment; but often, far too often, we get no dividend at all, or even worse, the bitter taste of ashes. Indeed, it has been said that the characteristic psychological problem of our time is the *dissatisfaction* that attends the fulfillment of our very most powerful desires (*op. cit.*, p. 531).

But these comments are in Feinberg's criticisms of psychological egoism. Do his examples not tell against anything like the reward-event theory? Do they not presume desires operating prior to, and independent of, any associated reward event? Not necessarily. For one thing, if these desires are really autonomous, why should dissatisfaction, or failure of satisfaction, which can attend their fulfillment, be a psychological problem at all? One might regret the absence of this "dividend," but motivation should carry on. Nevertheless, there is an important issue here. Its full discussion must wait for our third general topic—the direction of causation. For now, I shall make one crucial point. Consider the many cases where a person seems to pursue a goal that can only come to fruition after she is dead. The person never will experience that fruition, nor, a fortiori, any (possible) associated reward event. So how can such events be sustaining such motivation?

Any adequate theory of human motivation must, of course, take account of the enormous role played by our cognitive capacities. We should expect to find, in a charming phrase from the technical literature, the "encephalization of motivation." So, with respect to the just-mentioned problem cases, the story might go several ways. With respect to novel, future goals whose fruition one could expect to experience, one might well have beliefs, grounded in analogy with already experienced goals, concerning the possibility of appropriate reward events. One would expect to be satisfied. Not only is this in many cases at least introspectively implausible, however, it is also antithetical to the real spirit of RET. For on this hypothesis, a belief (granted, a belief about some future reward event) would be intrinsically motivating. Furthermore, this would not handle cases where one does not expect to experience the fruition of the goal.

---

nent is a pervasive loss of the capacity to experience pleasure—the technical term is anhedonia" (as reported in *The Times-Picayune*, New Orleans, September 2, 1988). If this anhedonia is not merely one more symptom, but the key to the loss of motivation, then we are close to RET.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Far more likely is the hypothesis that we have the ability to envisage future, or possible, or merely imaginary states of affairs, and that such envisagement, particularly when vivid, can link directly to the reward event. (Is that not what fantasy is about?) So, any associated motivation would be cycled through and dependent upon the reward event. Clearly this is an empirical issue. And one of the great advantages of finding a neurological basis of the reward event is that this could provide a way of gathering fairly strong evidence. It would be very difficult to determine introspectively whether one was being motivated solely by a present belief about some future state of affairs, or whether a present satisfaction arising from the evisaging of that state of affairs was the real motivational key. But if we could test neurologically for the presence or absence of the reward event, this would be important independent evidence. And it could go either way, so the more "Kantian" picture of motivation by belief alone could turn out to be right.[12]

A final comment on the sustaining of motivation, as displayed in long-term pursuit of goals. A full theory would need an adequate account of the nature and role of habit and whether it can operate in full autonomy from any reward events. But we should remember that other creatures condition very effectively with intermittent rewards rather than rewards on every trial; indeed, intermittent and unpredictable "schedules of reinforcement" produce the response patterns that are most resistant to extinction.

(2) *monism or pluralism.* How many ultimate objects of motivation? I have been sketching a monistic answer to that question—only one, the reward event. Since RET is intended as the sketch of an empirical theory, however, we should be flexible here. What even counts as one kind of event or object is highly relative to theory and interest. In the literature on ESB in the rat, there is both the emphasis on the phenomenon as a potentially unifying paradigm across other currently recognized forms of motivation, and, at the same time, there is a constant experimental search for different ESB locations that would be differentially associated with hunger or thirst or what have you.

In addition, it must be admitted that as the term 'motivation' is currently used in the psychological literature, it is seen as a pluralistic phenomenon. More accurately, there is a certain flexibility, perhaps even ambiguity, in what is being focused on. A brief sketch should clarify this point.

Studies of motivation generally recognize three interconnected phenomena. First, there are variable internal states of the organism

[12] Cf. fn. 13, alternatives (6) and (7).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that are postulated to account for variations in behavior, e.g., why a creature sometimes eats and sometimes does not in the presence of food, sometimes drinks and sometimes does not in the presence of water, etc. This internal condition (sometimes labeled "drive") usually gets the motivational label—hunger, thirst, etc. And, of course, such states need to be as pluralistic as the different kinds of behavior to be accounted for. Then there are the "goal objects"—food, water, etc.—which, plausibly enough, are seen as the objects of their respective motivational states. These are closely interdefined. Hunger *is* the motivational state that has food as its object. Food *is* the object of the motivational state of hunger. Sometimes, however, the goal object is defined in terms of some consummatory action performed on the object, so that the goal would be, e.g., eating the food. Finally, since much study of motivation has taken place in learning situations (e.g., operant conditioning), something is recognized as a reward (a reinforcement of learning). And many researchers (perhaps most in their official policy statements) would define such reward operationally or behaviorally, so that the behaviors of eating, drinking, stimulating the brain are the rewards. With that tight a circle of interdefinitions, reward really does not play much of an explanatory role. If eating the pellet is a reward because it increases the probability of the bar pressing that produces the pellet, then it seems less than illuminating to say that this learning, or "shaping" of behavior, takes place because it is rewarded.

But the reward-event theory is not committed to behaviorism. And much of the ESB literature really does seem to break out of the tight definitional circles. The reward event is not the *stimulating* of the brain (e.g., by bar pressing), it is the brain event *caused by* that stimulation, or so I shall read it. And this frees up a lot of explanatory space. It will no longer be true by definition that a reward event causes learning or shapes behavior. This will be a significant empirical claim concerning a contingent connection between certain brain events and certain patterns of behavior. And similar empirical anchoring would be provided for other interdefinitions. We might well start by saying that hunger is the internal state that makes eating food rewarding. But the task would then be to specify that internal state physiologically (there has been progress in this task). Now the claim would be that, when a creature is in thus and so physiological state, it will be selectively receptive to food stimuli, will eat food if available, and that this eating will trigger the reward event.

The immediate relevance of all of this is to note that, if something like this were correct, one might, for different purposes, wish to emphasize the differences among the physiological states (so there would be many different motives), or the differences among the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

external behaviors and objects (so there would be many different objects of motivation). I have no wish to legislate about that. But if physiological states become motives, and if behaviors or objects become objects of motivation, if and only if they become connected in the appropriate way with the reward event, then this is a kind of deep monism, which could be expressed by saying that there is only one basic motive and only one basic object of motivation. So whether the monism of RET is accurate or fruitful will depend on how the theories develop and what future research turns up.

As a postscript to this section, we might note what implications this kind of monism would have for the egoism issue. In one sense, one would still be able to make all the distinctions the anti-egoist wishes to make. Let us say that according to RET, only the reward event is ultimately or intrinsically rewarding or motivating. Many other behaviors, events, thoughts, or what have you would, however, either through a learning process or by being "wired in" that way, connect directly to the reward event. Let us say that when so connected (and this could change) they are then directly rewarding or motivating. (The contrast would be with those behaviors one would be motivated to perform only because they are seen as means to that which is directly rewarding.)[13] Now note that there could be two different

[13] The contrast is incomplete. Even if no external goal is ever intrinsically rewarding, or ever becomes functionally autonomous with respect to the reward event, we would still need to distinguish among:

(1) *innate direct goals:* external goals that are innately directly rewarding;
(2) *innate derivative goals:* external goals that, innately, are rewarding only through a "wired-in" association (contingent connection) with some external goal that is innately directly rewarding;
(3) *learned direct goals:* external goals that are not innately rewarding in either sense (1) or (2), but where experienced associations with goals of type (1), type (2), or (perhaps) type (4), produce new contingent connections in the brain such that the goal becomes directly rewarding, and at that point becomes functionally autonomous with respect to other external goals;
(4) *learned derivative goals:* external goals which are not innately rewarding in either sense (1) or (2), but which, by learned association with external goals of either type (1) or type (2), come to be pursued, but remain functionally dependent on the type (1) and type (2) goals;
(5) *cognitively-derived direct goals:* external goals that are not innately rewarding in either sense (1) or (2), but where cognitive development (new beliefs, knowledge, concepts, etc.), independent of any association with any other rewards, produces new contingent connections in the brain such that the goal becomes directly rewarding;
(6) *cognitively-derived derivative goals:* external goals that are not innately rewarding in either sense (1) or (2), but where learned beliefs about how that goal can lead to or help produce, ultimately, some external goal that is directly rewarding cause it to become a goal;
(7) *cognitively/affectively-derived derivative goals:* external goals that are not innately rewarding in either sense (1) or (2), but where learned beliefs

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized By Arya Samaj Foundation Chennai and eGangotri

individuals, one of whom finds the suffering of others directly rewarding while the other finds helping others directly rewarding. Deep down, they share reward-event motivation, but for social and moral purposes it is the differences that count. They are very different kinds of people, and of course that matters. But if the anti-egoist wants to claim that some people are ultimately or intrinsically motivated by, e.g., perception of the needs of others, then, if RET is correct, they are wrong about that.

(3) *direction of causation.* Is desire or satisfaction causally primary? Do we desire *X* because we have found it satisfying, or do we find it satisfying because we already desire it? What I want to do is to define the issue so that it really will be an empirical question.

At the level of our common-sense psychology, the phenomena seem a very mixed bag. Sometimes we experience something for the first time, find it enormously satisfying, and thereafter desire it. Sometimes we desire something, pursue it, and on obtaining it, do not find it satisfying. And sometimes we seem to find the obtaining of a goal satisfying precisely because we have wanted it so long and worked so hard for it. I shall leave it as an exercise for the reader how the reward-event theory would deal with each of these cases. (Hint: for the third case, there might be a good bit of not terribly effective self-deception. "*Surely* a goal I have worked for so long and so hard *must* be satisfying.")

Let us focus on just one fairly basic case—the desire for food and the satisfaction of eating—and use it to outline the deep empirical structure of RET. This is an appropriate case precisely because the scientific literature on which I have been drawing seems to answer the direction of causation question in the opposite way from RET. As already noted, hunger is often defined as the internal condition which makes food rewarding, and that does seem to give desire the causal priority. Some of this impression may be an artifact of

---

about how that goal can lead to or help produce, ultimately, some external goal that is directly rewarding, plus some thought of or envisagement of the directly rewarding goal, which thought or envisagement is itself directly rewarding, combine to cause something to become a goal [types (6) and (7) are the alternative possibilities discussed above.

(8) *non-cognitively-derived direct goals:* external goals that are not innately rewarding in either sense (1) or (2), but where contingent changes that are neither cognitive nor involve associations with any other rewards (rewiring through brain surgery?) produce new contingent connections in the brain such that the goal becomes directly rewarding.

Which of these possibilities has any actual instances is one of the things one hopes further empirical research will determine. Depending on the results, we would have rather different versions of RET, which can thus be seen to be a quite general kind of theory, which opens up a good deal of empirical space.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the official, operational definitions. But the issues are, or can be, empirical.

Let us suppose that we have identified the reward event neurologically. Let us also suppose that we have an independent, physiological identification of hunger, the internal condition which usually motivates a creature to eat and seems to make the eating rewarding. (I do not want any of this to be true by definition.) Now let us suppose that we can independently control the neurological event and the physiological state to produce the following experimental situations.

First, we determine that our creature is in the physiological state of hunger. We allow him to eat, but we suppress the neurological reward event. Let us suppose that the result is that the creature soon ceases trying to eat, seems no longer motivated to eat, even though the hunger condition continues or intensifies. (Have we modeled anorexia?)

Second, we determine that our creature is in a condition of satiety and thus ordinarily would not be motivated to eat. Now we artificially induce ingestion of food and, at the same time, trigger the reward event. Let us suppose that the result is that the creature now begins to eat and continues to eat as long as the reward event is triggered. (Clearly we are artificially bracketing any aversive events resulting from such overeating).

Such Mill's methods results would certainly suggest that the reward event, the satisfaction event, is causally primary. When the reward event is absent, there is no motivation; when it is present, there is. On the other hand, it might not work out that way. It might turn out that, in the absence of the reward event (case one), the creature is still motivated to eat, and that, in its presence (case two), it is not.

Let us consider those two very different possible results for our experiments, and focus on the reward event itself. We can now identify two very different theories concerning the role of reward or satisfaction with respect to motivation.

Before doing so, I should emphasize that I want the reward event to be defined solely in terms of satisfaction. I do not want it to be characterized as being intrinsically motivating by definition, or to have such intrinsic motivation be what it means for it to be intrinsically satisfying. I want it to be a question of contingent fact whether the intrinsically satisfying is intrinsically motivating. With that caution (which shows how deeply nonbehavioristic this approach has to be), we can sketch our theories.

First, there is the by-product theory, which would be consistent with the second set of results of our experiments. On this theory, creatures are intrinsically motivated to pursue a number of different

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

kinds of objects or to perform a number of different kinds of actions with respect to those objects. In many instances, though not all, when the object of motivation is obtained, this causes a reward event, but the occurrence of such events plays no, or at least no central role in the production or sustaining of motivation. Such reward events are mere by-products of the essential "machinery" that explains the creature's actions.

Then there is the reward-event theory, which would predict the first set of results. On this theory, the reward-event mechanism has evolved to play a central, functional role in the behavior of many organisms. Either through the evolutionary history of the organism or through its individual learning history, the reward event will have developed contingent connections with other physiological states, with perceived external events, with sundry cognitive states, and with various tendencies to action, in such a way that the reward event is the central causal key, an affective key, to why creatures do what they do when and where they do it.

And, of course, there is all sorts of middle ground. It could turn out that both the reward event and a number of other different kinds of objects or activities are intrinsically motivating. We may be a mixed bag deep down.

One of the most controversial aspects of this attempt to make the role of reward events (or felt satisfaction, or pleasure) a genuinely empirical issue is the insistence that any connection between the intrinsic satisfaction of the reward event and its power to be intrinsically motivating be a contingent connection. Can it even make sense to say that something is intrinsically satisfying but not at all motivating, not even in the minimal sense of wanting the experience to continue? Ironically, it is the more extreme forms of the by-product theory which require such a possibility. But I am on their side on this issue. I do not see any conceptual incoherence in imagining a creature that finds a particular experience totally delightful while it is going on, but experiences no regrets at its termination and is not at all motivated to bring it about again. Perhaps it would be some totally sedentary organism which must simply accept whatever the environment throws at it. It would, after all, be pointlessly frustrating to want to change things or to bring them about, if there were absolutely nothing one could do about it. And with a neurological ground for reward or satisfaction, we could even test for this independently of the standard behavioral, motive-driven tests for reward.

It is highly improbable that there are such creatures. My sense of the evidence is that reward or satisfaction does play a deeply functional role, and that it could well turn out to be so central that something like the reward-event theory is correct.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

I shall note just one rather major implication. If reward events (or affective states in general) do play a central functional role, and if they are distinct events or states only contingently connected with motivation and cognition, then that favorite model for explaining action, the belief-desire model, is seriously incomplete.

III

There are many questions left. If reward events have a central function, how widespread are they in the biological kingdom? Might the basis of them turn out to be sufficiently independent of our neurology, or even the rat's, so that simpler organisms could also be shown to operate on a reward-event system? How could one test for this without falling back into purely behavioral definitions of reward? If a reward-event system only appears in neurologically more complex creatures, what is the significance of that? Is such a system essential for the having of genuine desires, in the sense that the fulfilling or frustrating of those desires really matters to the creature? And what would be the significance of that with respect to value and morality?

But those questions must wait their turn. For now, let me summarize my sense of the probability of the reward-event theory with respect to our three issues. On the internal/external issue, it seems to me very likely that RET is correct, that some internal affective events play a central role. On the monism/pluralism issue, the evidence is still very much out. There are those straws in the wind suggesting deep neurological similarities underlying the great variety of externally identified rewards. But it could well turn out that, whereas internal affective events are crucial, the differences among them are theoretically more significant, so that we should speak of many different kinds of reward events. Finally, on the causal-direction issue, this is tied up with the question of the function of reward events. And here my sense is that such a system is likely to have evolved because it plays a key functional role.

One of my central aims has been to define the issues empirically, so we do not just settle them by definition. Although I am inclined at present to put my bets on the reward-event theory, with its internalism, monism, and causal primacy of satisfaction, I think we are very far from knowing enough to settle these questions concerning motivation, human or otherwise. The winds of science will blow where they may. In the meantime, we can be a bit more circumspect about what we put in our tribal folklore.

CAROLYN R. MORILLO

University of New Orleans

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

MUTUAL BENEVOLENCE AND THE THEORY OF HAPPINESS*

What do you want to do?" she said.
"Whatever you want to do."
"I want whatever's best for you."
"What's best for me is to please you," I said.
"I want to make you happy, Jack."
"I'm happy when I'm pleasing you."

White Noise, Don Delillo[1]

IMAGINE just two people, each of whom has no desire other than for the satisfaction of the other's desire. The object of the one's desire is the satisfaction of the other's, but the satisfaction of the other's desire would be constituted only by the satisfaction of the one's, and so on in a loop. The structure of this case is important for moral theory, by way of the havoc it wreaks on the notions of benevolence and happiness.[2] The rough structure of the argument I shall offer in support of this thesis may be summarized as follows. It may seem that the loop depends on the absence of any independent, first-order desires. Although the presence of independent desires supplies some content, it is not a general way out of the loop, however, since there is no reason to think that altruistic or benevolent desires will, in general, conveniently refrain from aiming at the satisfaction of the other's benevolent desires. As an example, this loop will be shown to occur under certain assumptions of Joseph Butler's moral psychology. The occurrence of the loop under the Butlerian assumptions shows that the loop can occur in realistic situations, but, moreover, it suggests that, without deep changes in the ordinary view of happiness, both benevolence and happiness may be empty in a disturbing way. The loop problem depends on, and so puts a certain pressure on, a satisfaction conception of happiness, according to which happiness is, in one way or another, the satisfaction of passions or desires. Butler himself may have held an interesting alternative

* Earlier versions of this paper were presented at various conferences, where I benefitted from discussions with John Broome, Nicholas Denyer, James Griffin, Shelley Kagan, and Patricia Werhane. I am grateful to the University of California/Irvine for a summer research fellowship in 1988. I also received helpful comments from Rachel Cohon, Tyler Cowen, Alon Harel, Michael Hudson-Smedley, Greg Kavka, Penny Maddy, Gary Watson, Joshua Weinstein, and Peter Woodruff.

[1] New York: Viking, 1985. Thanks to Andrew Cutrofello for pointing out this text.

[2] The notions of a person's good, happiness, and well-being will be used interchangeably.

0022-362X/90/8704/187–204

© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

conception of happiness which is not subject to the loop and its associated problems.

I. THE PRESENCE OF MORE PEOPLE IS NO WAY OUT

First, an obvious point. In the interest of realism, consider adding people to the two in the original example. The presence of, for example, a third person will obviously be irrelevant if the first two remain as originally described: interested only in each other. The point of introducing the third person could only be that it may provide a way out of that two-person loop. So suppose that at least one of the original two people is concerned for the third person. We should suppose this third person is like the other two, however, in desiring nothing but the satisfaction of some other person's desires. But the only others, the original two, desire only the satisfaction of one of the two others as well: either the satisfaction of the other member of the original pair, or of the new third person. In either case, every individual's desire is involved in a loop. Clearly, the same holds for four people, or any number.[3]

If the number of individuals is infinite, the problem is, in a way, even simpler. What is puzzling about the loop, after all, is that tracing it is an infinite process even with only a finite number of people. If there are infinitely many people, and the desires are set up so that a loop never forms, the process of tracing the desires to their objects is still infinite.[4]

Obviously, then, the emptiness of the original loop can be present in cases involving any number of people.

II. DOES BENEVOLENCE FADE OUT?

The thought of such a desire chain involving a large number of people naturally prompts the observation that we typically care less about the loved ones of our loved ones than we do about our loved ones themselves. I may want your mother's surgery to be successful because of my concern for you, but if that is the ground of my concern for her, it will be, in some sense, weaker than my concern for you. If chain-linked desires of this sort do diminish in strength, then, if the chain is long enough (e.g., your mother's friend's surgery), the desire strength might be thought to fade out. If this happened, it

[3] Consider person *A*'s desire for the satisfaction of some other person's desire, in a group with n members. Even if the chain of desires gets to the nth person without looping, if n is a finite number the nth person's desire will have to be for the satisfaction of the desire of one of the others (*A* through n), and that will close the loop.

[4] We cannot just ignore this possibility on the grounds that there are never infinitely many people alive. That would be to assume without argument that the desires in question cannot concern future persons, of whom there may or may not be infinitely many.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

could be argued that there is no real infinity involved in the original desire. A closer look reveals, however, that this sort of fade-out is irrelevant to the occurrence of a loop.

It is the strength of the desires that might be thought to fade out, but strength is beside the point. To illustrate, it will be useful to start with a case where the desires do not form the sort of loop in question; so suppose that Joe wants only that Sam get whatever he wants, and Sam wants only to get a job. If we say that this shows that Joe wants Sam to get a job, then the question might arise whether Joe wants this less strongly than he wants Sam to get whatever he wants. But the problem concerning us does not require us to state things this way. Rather than ask what further desires Joe has as a result of the chain, we need only ask: What state of affairs would satisfy the original desire of Joe's, that Sam get whatever he wants? The answer in this case is that Sam's getting a job would satisfy it. But now how can the fade-out thesis be stated? Could it be held that "Sam's getting the job" would satisfy Joe's benevolent desire less than "Sam's getting what he wants"? Of course not; either one satisfies it completely.

It is common practice to suppose that our attitudes toward certain events can vary with the description of the event. For example, there may be one event that is both the setting down of a piano one is helping to move, and the crushing of the neighbor's cat. The event may be desired under one description, and abhorred under the other description. The attitudes need not vary so greatly with the different descriptions; one may care greatly about an event under one description, but hardly at all under another. So consider an event that is both the satisfaction of Sam's desire, and the getting of a job by Sam. The fact that Joe may care less about the event under the latter description than under the former does not impugn the strength of Joe's benevolent desire for that event. If we change Sam's desire for a job to a desire for the satisfaction of the desire of some third person, say, his mother, this only lengthens the chain that determines what event would satisfy Joe's desire. The fact that a new description comes to apply to the event (such as, 'the satisfaction of Sam's mother's desire'), a description under which Joe cares little about it, does not diminish Joe's concern about this event, since, whoever else it may satisfy, it satisfies his friend Sam.[5]

[5] Joe will typically care less about the event that would satisfy Sam where it is described in any of the ways Sam cares about it. And if that second description itself involves the satisfaction of a third person (and Joe had at least some concern for it under this description), he will care even less about it under the description that matters to the third person, and so on. (Of course, Joe may coincidentally have some separate strong concern for the event under one of these further descriptions, e.g., for the welfare of some person whom these others happen to care about, too.)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

If Joe's concern does not fade out in the relevant sense where Sam's only desire is for his mother's satisfaction, then, for exactly the same reason, it will not fade out where Sam's only desire is for Joe's getting what he wants, a desire which gives rise to a loop.

III. INDEPENDENT DESIRES

Suppose there are not only the desires about each other, but also other "independent" desires whose object is not the satisfaction of anyone's desires. Consider the case of just two people again. Each still has the desire for the satisfaction of the other's desire (call this the benevolent desire), but now each also has another desire as well. Suppose *A* desires to eat an apple and *B* desires to eat a banana, while each still wants the satisfaction of the other's desire. If the benevolent desires in question are just for the satisfaction of *any* desire of the other, then the puzzle does not arise. For example, *A*'s benevolent desire for the satisfaction of any desire of *B*'s is satisfied by the satisfaction of *B*'s banana desire. It does not require the satisfaction of *B*'s benevolent desire, and so there is no loop.[6] If *A* gets an apple and *B* gets a banana, then all the desires in the example are satisfied, including the benevolent ones.

But here the benevolent desires are conveniently undemanding. The presence of independent desires is not a general way out of the loop at all. Consider, for example, a benevolent desire of *A*'s that aims solely and specifically at the satisfaction of *B*'s benevolent desire, and vice versa. Clearly the mere presence of the other, independent desires does not prevent the loop. Admittedly, the example where *A*'s benevolence aims solely and specifically at the satisfaction of *B*'s benevolence is contrived, but then so is any example in which the object of *A*'s benevolent desire is held specifically to *exclude* the benevolent desires of *B*—those desires which can result in a loop. The introduction of independent desires can only avoid the loop in contrived examples of this latter sort, since as long as the satisfaction of another's benevolence is part of the object of everyone's benevolence the interminable satisfaction loop will remain.

Here is the case that should concern us. Suppose *A* has a benevolent desire that aims at some level of satisfaction of all of *B*'s desires, and *B* has the same with regard to *A*. This is not just an example

[6] John Rawls suggests this sort of approach, but it undermines his own use of the loop case, which is to defend mutual disinterestedness among the parties in the original position. Since the parties will have some independent desires whether or not they have benevolent ones, the fact that mutual "*perfect* altruism" (nothing but "dependent" desires) is empty is no grounds for assuming the absence of dependent (including benevolent) desires altogether. The criticism of this solution below would seem better to support that Rawlsian move. *A Theory of Justice* (Cambridge: Harvard, 1971), p. 189.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

concocted to yield a loop; it is, in effect,[7] Butler's view of benevolence and happiness. He never discussed the puzzle to which this appears to give rise, namely, that the satisfaction of *A*'s benevolent desire would require the satisfaction of *B*'s, which in turn requires the satisfaction of *A*'s, and so on. But if Butler's account is correct in certain respects, then there are real-world loops, and the emptiness inherent in them is more than a curiosity (but more on that below).

IV. BUTLER ON BENEVOLENCE AND WELL-BEING

On Butler's view, the psyche is populated by multiple passions. A passion is satisfied when its object is obtained.[8] The object of the passion of hunger, for example, is food. One of the passions that all people have by nature is benevolence, which has as its object the happiness of others. Another is self-love, which has as its object one's own happiness.[9] Happiness does not require the full satisfaction of

[7] Substituting passions for desires, as explained below, yields Butler's apparent view. Note, however, the interpretive questions raised in section VII.

[8] Joseph Butler, "Fifteen Sermons Preached at Rolls Chapel," (excerpts) in *British Moralists*, D. D. Raphael, ed. (New York: Oxford, 1969). The term 'passion' in Butler is more complex than the notion of desire as commonly understood today. For example, we all have the passion of hunger as a part of our psyche, yet we are only hungry intermittently. Roughly, the passion for food is a category of potential desire for food. As such, it seems wrong to say that passions actually have objects even when they involve no actual desire, but a few brief distinctions will show how Butlerian passions do have objects.

The relation of passion to desire is similar to the relation between having a concept and having a thought or belief. The concept is a category of potential thought or belief, and endures between episodes of these. Satisfaction of a passion is really satisfaction of a desire which is the actualization of a certain passion. We may speak of Butlerian passions as being in one of two states: *latent* or *actual*. A passion is actual rather than latent by virtue of the existence of a desire of a certain kind, i.e., one having an object that falls within the category defined by the passion. It is that category that gives the object of the passion itself.

We can also speak of two different phases of desire: *background desire*, and *effective desire*. This is a different distinction from that between latent and actual passions. Effective desires, let us say, are desires that are part of the rational explanation of some actual behavior. We have desires even when they are not effective in this sense. For example, I want to have dinner sometime tonight, but my wanting this does not figure in explaining any of my behavior during most of the day. The desire is present, but not effective. It is not irrelevant since attributing it supports certain counterfactual propositions in an economical way (e.g., if I were given the choice now of later having dinner or no dinner, I would choose dinner unless paid in some way to forego it, etc.). Both background and effective desires have determinate objects, unlike latent passions. The relevant latent passion in the dinner example is hunger. It is the category of potential desire for food. The passion of hunger is nonlatent whenever one has a desire, either background or effective, for food of some specific kind, or at some specific time. That is, a passion becomes nonlatent by acquiring a determinate object, and both background and effective desires have determinate objects. A passion or desire has a determinate object when there is some state of affairs that would satisfy it. Latent passions, then, have objects in a certain sense, but, unlike both background and effective desires, they do not have determinate objects.

[9] The puzzle might seem to appear here in even a more pure form. Self-love is a passion, however, that has as its object a balanced satisfaction of one's own *particu-*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

all passions; it is rather a balanced satisfaction of all of the passions, with the requisite balance determined by nature.[10]

Of course, there would clearly be a loop if everyone's only passion were benevolence, but is the loop avoided by the presence of the other passions? We have already seen (sect. III) that, unless benevolence is specified as excluding from its object the benevolent passions of others, the addition of independent passions will not avoid the loop.

It is contrived and implausible to stipulate, however, that benevolence excludes from its object the benevolent concerns of the cared-for others without any reason to think they are any less important to that person's well-being. Butler indirectly gives good reasons for thinking that such a restricted benevolent passion could not be regarded as having as its object the happiness of others; its object would no longer count as happiness. One of Butler's recurrent lessons is that attention to one's benevolent passions is integral to happiness; it is an ineliminable part of the balanced pattern of satisfactions which is, according to nature, human happiness. It is precisely the neglect of one's own benevolent passions which is the premier failing of self-love, and the result of the failure is selfishness and lost happiness.[11] So, unless the satisfaction of your benevolence

---

lar passions. The particular passions are just all those except self-love, and if self-love were not excluded in this way the puzzle would indeed arise. Self-love would have as part of its object its own satisfaction. It is an interesting question whether it is legitimate to specify the object in this way (as Butler does), while doing so is here criticized as contrived in the case of benevolence. I do not believe the same arguments are available in this same-person case. Whereas the satisfaction of a loved one's benevolence could naturally be thought to be a crucial component in their well-being, I do not see how the same could be said about the satisfaction of one's own self-love, and, if not, then it is appropriate to exclude it from its own concern. One cares about the objects of one's passions, and one cares about one's own happiness (which involves the satisfaction of those passions), but why say one *also* cares about the satisfaction of one's own concern for one's own happiness?

[10] I offer the following textual composite, from Butler, as some support for this interpretation:

> Every bias, instinct, propension within, is a real part of our nature but not the whole (§404).
>
> . . . Neither can any human creature be said to act conformably to his constitution of nature unless he allows to [conscience] the absolute authority which is due to it (§379).
>
> Happiness or satisfaction consists only in the enjoyment of these objects which are by nature suited to our several particular appetites, passions, and affections (§417).

[11] ". . . the greatest satisfactions to ourselves depend upon our having benevolence in a due degree." Butler, Sermon 1, §388 (references are to sections rather than pages, since these are used in the table of contents and index of that edition).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

is part of the object of my passion, my passion does not have your happiness as its object. But then it would not be the passion in question—benevolence.

Suppose that instead of wholly disregarding the benevolence of others, *A*'s benevolence ignores only the *A*-regarding desires of others, thereby still heading off any possible loop. The restriction is ambiguous. Does it mean that *A*'s benevolence does not seek the satisfaction of (1) only *directly* *A*-regarding concerns, or (2) any concern which directly *or indirectly* is *A*-regarding, as in *B*'s concern for *C*'s concern for *A*? If (1), then only two-person versions of the loop are precluded, and so the puzzle is not avoided. Cases like the following would still be allowed: *A*'s concern for *B*'s concern for *C*'s concern for *A*'s benevolent passion. If the proposal means to rule out directly *and* indirectly *A*-regarding passions as in (2), then the proposal must be rejected as *ad hoc* and unrelated to the real nature of benevolence.[12] After all, there is no distinctive content of a desire that will make it *A*-regarding. Any other-regarding desire may be indirectly *A*-regarding if the facts about other people's desires turn out a certain way. It is implausible to build into the passion of benevolence a clause which excludes any desires of others which, in the facts of the case, turn out to be *A*-regarding. This would mean taking *A*'s concern for the satisfaction of *B*'s benevolent passions to be automatically cancelled if some direct or indirect object of *B*'s benevolence happens to care for *A*. Such factors are irrelevant to the nature of *A*'s concern for *B*.

It is unacceptable, then, to try to avoid the loop by conceiving the benevolent passions as excluding from their objects either the benevolent passions of others in general or those particular benevolent desires of others which concern oneself.

V. WHAT IS THE PROBLEM?

If Butler's moral psychology is correct in key respects, then desire loops are a serious problem. It would, of course, follow that such loops can occur in realistic situations rather than merely in philosophers' concocted examples, and though that by itself would be of

---

This passage is ambiguous, since it attends to both the having and the satisfying of benevolence. The choice between these is important to our topic and will be discussed below.

[12] It is *ad hoc* if it is ruled out just because of the loop. It would be less *ad hoc* if the objection is that such a desire turns out not to be benevolent but self-interested. The fact that its satisfaction would benefit oneself, however, is not a sufficient ground for denying that it is benevolent. It benefits oneself only because the happiness of the cared-for person depends on one's own happiness, not because of any concern for oneself.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

some importance, it would not obviously be a problem. We could simply allow that such cases can occur in the real world and that, where they do, there are no conditions that would satisfy the passions involved. This would seem to shrink the significance of the loop problem, by analogy with loops in other contexts.

For example, it is well-known that similar loops can emerge in certain sets of statements, such as the following: *A* says only:

"*B*'s statements are always correct."

And *B* says only:

"*A*'s statements are always correct."

This statement of *A*'s is correct if and only if it is correct, and incorrect if and only if incorrect, and nothing else bears on its correctness. Such loops undeniably can occur in the real world, and so there are sentences for which there are no independent truth conditions. Assuming that some precise account can be given of this condition of "ungroundedness," as it has come to be called, the fact that it can occur may not present any special philosophical problems.[13] Similarly, the fact, important though it may be, that desires can be ungrounded in the real world, would not by itself present serious philosophical difficulties so far as I can see. At least that is not the sort of problem I want to allege.

There is more at stake in the Butlerian account of benevolence, however, than just the question whether desires *can* be ungrounded in real-world cases. If certain features of that view are correct, then benevolence is *always* ungrounded—nothing would count as satisfying anyone's benevolent desires. Most problematically of all, if Butler is correct about the relation between benevolence and happiness, then the general ungroundedness of benevolence renders human happiness utterly undefined. The conclusion is not just that happiness is difficult or even impossible to obtain; it is rather that there is, in principle, no condition that would count.[14] Now these conclusions

[13] For recent advances in the provision of such an account, see Robert Martin and Peter Woodruff, "On Representing 'True-in-*L*' in L," *Philosophia*, V (1975): 213–7; and Saul Kripke, "Outline of a Theory of Truth," this JOURNAL, LXXII, 19 (November 6, 1975): 690–716. Both are reprinted in Robert Martin, ed., *Recent Essays on Truth and the Liar Paradox* (New York: Oxford, 1984).

[14] On the Butlerian assumptions, self-love and benevolence are not just risky as Kripke says we must allow sentences involving truth to be. ". . . an adequate theory must allow our statements involving the notion of truth to be *risky:* they risk being paradoxical [or ungrounded] if the empirical facts are extremely (and unexpectedly) unfavorable." Kripke, *op. cit.* The present passage occurs on p. 55 in the Martin volume. In the Butlerian case, there is not just the risk but the guarantee of ungroundedness.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

do not render the theory incoherent; as we have said, there is no incoherence or contradiction in supposing that the emptiness associated with ungrounded desires can actually obtain. The conclusion that the very notions of benevolence and happiness are empty in this way is, however, surely radical enough to warrant a closer examination of the Butlerian claims that would jointly establish it. In what follows, the argument will be that, barring a radical change in our common sense views about happiness, the features of Butler's theory that conspire to enervate benevolence and happiness are well founded. The philosophical difficulty this presents is acute, but it can be made clearer only after a closer critical look at the key Butlerian positions.

VI. THE SUFFICIENT BUTLERIAN POSITIONS

The two features of Butler's theory that would void benevolence of any satisfaction conditions are:

1. Benevolence aims at the good (happiness) of others.
2. A person's good (happiness) necessarily includes some degree of satisfaction of their benevolent passions.

It follows from these that each person's good depends on someone else's good, giving rise to an infinite chain of dependence with either infinitely many individuals as links, or (more likely) a loop.[15] Benevolence is, therefore, ungrounded (in the logicians' term), leaving happiness unachievable in principle. It is not just perfect happiness that is affected, but happiness of any degree, as will be discussed below.

The first claim is stipulative. There could certainly be disagreements about whether real benevolence, or any passion, has the good of others as its object, but these will be beside the point as long as (2) can be defended while understanding benevolence in the way set out in (1).

Consider next claim (2). It says that a person's good involves some degree of satisfaction of the benevolent passions. Now, since the benevolent passions have as their object the good of some others, then it follows that each person's good depends on the good of some others. But this could easily seem too strong. Perhaps *A*'s benevolence can be satisfied, at least to some extent [and no more is explic-

[15] Without also assuming that *everyone* has benevolent passions it might seem that only those who have benevolent passions will be affected by the loop. (2) says, however, that everyone's happiness depends on benevolent satisfactions, and it does not restrict this to those who have benevolence. Therefore, (1) and (2) are sufficient to implicate everyone in benevolence's ungroundedness, since there is nothing that would count as the satisfaction of benevolence for anyone—whether they happen to be benevolent or not. Still, the universal presence of benevolent passions is one natural way of supporting the connection between benevolence and happiness in (2).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

itly required by (2)], by something less than *B*'s actual happiness, or good life, such as, for example, some contribution to it.[16] In that case, one person's happiness would not in general depend on another person's happiness, and still benevolent satisfactions (of this partial kind at least) would be acknowledged as necessary for happiness. Whereas benevolence would still have the *actual* happiness of others as its object (unlike the earlier contrivance of excluding benevolence satisfaction of others from benevolence's object), happiness would not require that one's benevolence be satisfied in this "whole" way; mere contribution satisfaction would be sufficient for happiness. This would avoid any loop that would render benevolence ungrounded. Even on this view, *A*'s benevolence could not be satisfied in a *whole* way without bringing some *B*'s benevolence into the picture (though, as just noted, *A*'s happiness would not require this). But no one else's benevolence (neither *A*'s nor some other *C*'s) need enter since *B*'s happiness (which is the "whole" object of *A*'s benevolence) requires only a contribution satisfaction of *B*'s benevolence. That could be achieved without any benevolent satisfactions of any others, and so the chain stops short of a loop. This account of the relation between benevolence and happiness purports to ground benevolence, and save happiness from meaninglessness.

The above account, however, oddly leaves benevolence's primary object, the happiness of others, out of the pattern that constitutes the benevolent person's happiness. All that matter are contributions to that condition. It is important to notice that the proposed account does not just say that happiness does not require complete satisfaction of benevolence; it says that happiness does not require even a single instance of the object of benevolence, the happiness of others. All that is said to be required is something else: contributions to happiness. Mere contributions to the object of a passion are not generally satisfaction of that passion at all. For example, the passion for shelter is not at all satisfied by the possession of some nails.

The proposal can be clarified by giving a separate status to the passion that has as its object contributions to the happiness of others. Call it *contribution benevolence*. Let us grant that its satisfaction is a necessary component of happiness. But certainly contributions are not the only objects of benevolence; there is also the happiness of others. That object is sufficiently different from mere contributions to happiness to individuate a separate passion—call it *happiness benevolence*. We must now ask, however, whether satisfaction of *this*

[16] I do not mean an act of contributing to it, but a contributing part or component itself.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

passion is a necessary component of happiness.[17] If not, then the proposal succeeds at avoiding the loop. If it is, then happiness benevolence is ungrounded, and happiness is empty.

One could try to deny that the actual happiness of others is a necessary component of happiness, while still retaining the general view of happiness as the balanced satisfaction of the passions, by denying the existence of any fundamental human passion that aims specifically at the happiness of others rather than aim merely at contributions to it. That is, it could be argued that contribution benevolence is, at most, the only benevolence that is central and universal enough to be regarded as a necessary component of human happiness. And, as we have seen, that would avoid the loop. This move is inadequate, however, especially in the context of benevolence among loved ones. The benevolence of a parent for a child, or of one spouse for another, is more than a passion for contributions to their well-being; it is also a passion precisely for that well-being. Of course, one also desires lesser things such as benefits, and contributions to the well-being of loved ones. But nothing is more natural than to desire the well-being of loved ones itself; it is one of the fundamental human passions.[18]

In a cautious mood, we might wish to avoid such a broad appeal to human nature. Perhaps the claim does not hold for individuals in all cultures, or in all historical and material circumstances. Still, as long as there are some significant cultural or other domains in which the generalization appears to hold, then in that context the loop is guaranteed.

As another form of caution, we might note that, if the postulation of benevolence holds for almost all, but *not all*, individuals, even in a specific limited domain, that is enough to remove the guarantee of a loop. This is correct, but the situation would still be qualitatively different from the mere risk of a loop. Each person's happiness and benevolence would probably, even if not certainly, be rendered empty by loops. Whether these more cautious estimations are required or not, the scope of the problem remains considerable. For

[17] Treating it as a separate passion is only heuristic. We could just as well keep it as one passion and ask whether something more than contribution-satisfaction of this passion is required for happiness, e.g., "whole" satisfaction: the actual happiness of some others.

[18] If happiness were *defined* as requiring, among other things, the happiness of some other people, then the definitions would apparently be viciously circular. But this is avoided if happiness is defined simply as a balanced satisfaction of the passions, and each passion's role in happiness is regarded not as definitionally necessary, but as contingent on its actually being one of the human passions.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

simplicity I shall go on to state the case in the less cautious way, as though all humans had such benevolent passions.

While accepting that everyone has such a fundamental passion for the happiness of some others, one might deny that happiness requires its satisfaction. Suppose someone had everything else that matters, even the near happiness of those about whom one especially cared. If the other satisfactions were had in sufficient abundance, could the person not be happy? The question is whether this absence of benevolence satisfaction could be compensated by extra satisfactions in other areas. If it could, then the charge that the loop is guaranteed to occur for all people should be dismissed. One's happiness would not in general depend on the happiness of someone else; nor, therefore, would the satisfaction of one's benevolence depend on that of another. It is a third condition of the ubiquity of benevolence loops that satisfactions of other passions cannot compensate for the nonsatisfaction of happiness benevolence, that different satisfactions are not interchangeable in this way.

Allowing such interchangeability is, in principle, a way of avoiding the loop. Should it be allowed? First, probably few would believe (these days) that happiness is sufficiently homogeneous for there to be *no* category of satisfaction that is indispensable. Most have rejected the view of happiness as a single psychic quantity that is produced by such different activities as pushpin and poetry. Therefore, the lack of one kind of satisfaction may be irremediable via other kinds. If the rejection of interchangeability of satisfactions is based on the difficulty in specifying some single feeling or psychic state that is common to all satisfactions, however, then the point does not carry over to the sort of nonpsychological satisfaction in question here (where a desire for x is satisfied if and only if x). But there is an influential and structurally similar move that is made in the latter case. The view is that happiness is not some separate state (psychological or not) that is produced by each of the multiplicity of categories of satisfaction. It is argued to be, in John Rawls's phrase, an *inclusive* rather than a *dominant* end. It consists of those other satisfactions (*op. cit.*, §83). Such a view must deny general interchangeability in order to avoid positing a single dominant end after all, namely, *satisfaction.*[19]

[19] Views that allow general interchangeability of satisfactions could avoid the loop problem, but I believe such views are seriously flawed. That cannot be argued here, but if they are then it matters much less whether they can avoid the benevolence loop or not.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

It would not be surprising if some satisfactions necessary for happiness were indispensable in this way. But the indispensability may be limited to a few certain passions. We need to ask whether the set of indispensable satisfactions includes those of happiness benevolence.

The denial of complete interchangeability of different kinds of satisfactions is powerful; we no longer expect any satisfactions to be interchangeable. Perhaps some are, for special reasons or in certain circumstances. For example, it may well be that satisfactions can be interchanged in the context of a mediocre life. A mediocre life without achievement can be just as good as one with achievement if the former has some satisfactions that the other lacks. Mediocre lives are fundamentally flawed, and the flaws might as well be in one category as another as far as the degree of well-being is concerned. But a truly happy life, human flourishing, is not fundamentally flawed. That does not mean that only a perfect life is a truly happy life; even the happy life could always be better, and so there is no perfect life. Rather, what must be meant, at least in part, by a life that is not fundamentally flawed is one that is not deficient (at the very least, not wholly lacking) in respect of any fundamental component of well-being. No catalogue of fundamental components is needed here. Only one candidate matters for our purposes. Even if one has more of some other satisfaction than usual, a person whose passion for the happiness of others (one's happiness benevolence) is not satisfied, at least in the case of some loved ones, is deprived in a fundamental way. Of course, such a life could still be valuable, but a truly happy life would not have this fundamental flaw (or so a satisfaction account of happiness seems to be compelled to say.)[20] I conclude that satisfactions are not interchangeable in the way that would be needed to avoid the loop.

The importance of the loop problem would be vastly greater if it affected not only satisfaction views that are concerned with *actual* desires or passions, but also with the currently more popular satisfaction views that deal with desires that are, in certain ways, *ideal*. The latter sort of satisfaction view is widely preferred for avoiding the problem of rash, ill-considered desires. Still, such views may well be subject to the same kind of benevolence loop as actual desire theories, and so to whatever problems this might involve. The following considerations lend some support to this possibility.

[20] Rawls apparently sees benevolence as an indispensable part of a person's good, §75). This does not yet make benevolent *satisfactions* indispensable, which is the issue. There seems no reason to doubt it, however, as will be suggested below.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Many adopt some version of the view that what is good for a person is what they would desire if they were fully informed, and I shall concentrate on these.[21] Suppose that

1. One thing anyone would want if fully informed is the good of some others.
2. A (nonhypothetical) person's good requires that their hypothetical benevolent desires [from (1)] be satisfied.

Clearly then, one person's good would depend on another's, in a loop.

Do the "ideal theorists" need to accept (1)? The fully informed person would surely want a life with deep personal relations of love and friendship.[22] These in turn would involve desires for the good of others. But so far all this shows is that *having these desires* is part of the good life. It does not yet show that the good of others (the object of the desires) is part of the good life as (1) implies.

The only way to avoid the conclusion, however, is to suppose that the fully informed person would *want to want* the good of others—would want to have such concerns, but would not necessarily *want the good of others*. There is no contradiction in this, but nor is there any good reason to believe it. Without some special explanation, it is natural to assume (in the context of a satisfaction account of happiness) that, if having the passion is part of one's good, then, for that reason, so is the object of the passion. This follows from the normal efficacy of passions. But even apart from that point we would expect the fully informed to want the good of at least some others in the more direct way.

Clause (2) prompts a distinction between a person's benefit, and their happiness or good, since no particular hypothetical-desire satisfaction is necessary for a person's benefit. But are there any that are necessary for one's happiness as (2) states? The same arguments used above (for the actual-desire version) for the necessity of some satis-

[21] Consider Rawls: "Thus the best plan for an individual is the one that he would adopt if he possessed full information. It is the objectively rational plan for him and determines his real good," p. 417). John Harsanyi's view is that "true" preferences are those one "*would* have if he had all the relevant factual information" and was rational. Social utility should be based on these hypothetical "true" preferences, though the account of individual utility should also consult actual "manifest" preferences "in a suitable way"; see "Morality and the Theory of Rational Behaviour," in *Utilitarianism and Beyond*, Amartya Sen and Bernard Williams, eds. (New York: Cambridge, 1982), pp. 39–62, esp. pp. 55–6. James Griffin has a similar view: " 'utility' is the fulfillment of informed desires," "an 'informed' desire is one formed by appreciation of the nature of the object"; cf. *Well-Being: Its Meaning, Measurement, and Moral Importance* (New York: Oxford, 1986), p. 14.

[22] Rawls, §75, and Griffin, pp. 67–8.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

faction of happiness benevolence can apparently be applied in the ideal-desire theories as well. The ideal-desire theories should not diverge from actual-desire theories in the case of actual desires that represent deep, central human passions.

Ideal-desire satisfaction theories of happiness seem compelled to grant (1) and (2), and that would give rise to the same loop that attaches to actual-desire satisfaction theories. Therefore, there is reason to doubt that moving from actual- to ideal-desire accounts of well-being will give a decisive advantage in avoiding the loop.

If our theory is that a person's happiness depends on satisfactions of desires, passions, etc., then it would be hard not to allow that the happiness of at least some people could, in principle, depend on the happiness of some others. But if this is allowed in principle, then all that is needed for the problem of the loop to emerge is a certain empirical situation, namely, that this situation which is allowed by the theory to hold for at least some people happens to hold for all people. And from within the view that happiness is constituted by satisfactions of (real or ideal) desires, it would apparently be arbitrary not to include as necessary the satisfaction, to some degree or other, of as central a passion as that for the happiness of some others. The problem is that including these passions will subject happiness to the consequences of a satisfaction loop, consequences which are severe: the ungroundedness of (happiness) benevolence, and the emptiness of happiness. This outcome is not incoherent, but still, it is such a radical conclusion (and such an undesirable one) that the assumptions that lead to it deserve scrutiny, even skepticism. They are, at the very least, put under pressure.

VII. IMPLICATIONS

If the relevance to one person's happiness of the happiness of others cannot be denied within a satisfaction model of happiness, we must ask on what deeper assumptions the radical conclusion depends, and whether they should be retained under this pressure. The satisfaction model of happiness should itself be scrutinized in this context. Now, although the radical conclusion can be easily shown to depend on a satisfaction model, that is insufficient reason for abandoning the model. We might, on balance, have more reason for retaining the satisfaction model than for fleeing the radical conclusions about benevolence and happiness. But that is what needs to be determined.

No full evaluation of the satisfaction model is possible here, but the importance and interest of such a project are heightened by two considerations which can be briefly discussed. First, and most importantly, the loop problem can be shown to depend on a satisfaction model of happiness. Second, since the problem has emerged out of a

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

consideration of certain views of Butler's, it is of some interest to note that the issue of whether happiness is a matter of satisfaction or not is present, though not explicit, in Butler's own thought.

First, the loop cannot occur without the assumption that a person's good depends on the satisfaction of passions or desires in some way. Suppose instead that happiness consisted in *having* certain passions (to certain degrees and in certain combinations), apart from whether they were satisfied. One person's benevolence would still have another person's good as its object, and that second person's good would still involve their benevolence in a certain way. It would not be the satisfaction of that benevolence, however, but only the having of it that figures in the person's good. Therefore, having someone's good as the object of my benevolence would not involve also taking on the objects of that other person's benevolence. Their good is comprised of their character, and their having the proper character is largely independent of their getting what they, in virtue of their character, want. The point applies more generally as well. Whatever happiness does depend on (character, pleasure, an objective list of goods, etc.), as long as it does not depend on the satisfaction of passions and desires my benevolence will not automatically take on the object of yours. A nonsatisfaction account of well-being, then, would avoid the puzzle of ungrounded desire as it arises in the case of mutual benevolence.

There is one kind of satisfaction view that is well positioned to avoid the loop. Rather than deny the relevance of the happiness of others in an *ad hoc* way, one might deny the relevance to one's own happiness of any state of affairs that is not actually an effect on oneself. Indeed, I think there is much to be said for such an "effect requirement." If it is admitted, then the only acceptable satisfaction view of happiness will be the view that

> something bears on one's happiness if and only if it both satisfies some (real or ideal) desire, and is an effect or change in oneself—that is, if and only if it satisfies some desire for some effect on oneself.[23]

Like any view that denies that the happiness of one person requires the happiness of some others, this view avoids the loop. I mention it only because the device by which it avoids the loop is well motivated in its own right. But, of course, it is forced to bite the bullet and disqualify many kinds of desire satisfaction that more standard satisfaction views would like to include. One could, of course, accept the effect requirement without trying to reconcile it with a satisfaction

[23] That need not be the description under which the person desires it.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

view of happiness at all, three of the salient alternatives being character views, hedonist views, and objective-list views.

Although the language of satisfaction and gratification occurs often in Butler, there is some reason to wonder whether he might have held a character view of happiness rather than a satisfaction view. First, while it may seem that a central Butlerian argument, the argument against psychological egoism, depends on a satisfaction view, this is not actually so. He argued from the assumption that a person's good consists in the balanced satisfaction of one's passions, that therefore, if one's only passion were for one's own good, it would simply have its own satisfaction as its object (*op. cit.*, sect. 383; cf. also fn. 9). This was held to be untenable, and we can see why in the fact that it is an instance of the sort of ungrounded desire that is our central topic; indeed, it is the purest case since it involves only one person and one desire. Butler concluded that there must be passions for things other than our own well-being. It may look, then, as though it is part of the Butlerian rejection of psychological egoism to suppose that our well-being consists in getting what we care about in certain ways. But Butler's deeper point does not depend on this assumption. Butler's lesson is that we cannot understand a person's good in isolation from what matters to the person besides her own good. This, if true, would be enough to refute psychological egoism. I have stated this lesson carefully, however, so as to show that it does not require a satisfaction view of well-being. The fact that an account of well-being is impossible without discussing what else matters to the person does not show that well-being consists in getting what matters to us. It could just as well consist in the mattering itself. Butler's lesson is consistent with holding that well-being is a pattern of concerns rather than a pattern of satisfactions.

There is also more positive evidence that Butler's own views betray at least some ambivalence about the choice between a character view and a satisfaction view of well-being. In one of the more ambivalent formulations, he writes, "the greatest *satisfactions* to ourselves depend upon our *having* benevolence in a due degree" (*op. cit.*, sect. 388; emphasis added). We see elsewhere, however, that Butler sometimes speaks of a kind of satisfaction that is not the obtaining of the object of any passion. For example,

> Happiness consists in the gratification of certain affections, appetites, passions, with objects which are by nature adapted to them. . . . Love of our neighbour is one of those affections. This, considered as a *virtuous principle*, is gratified by a consciousness of *endeavoring* to promote the good of others; but considered as a natural affection, its gratification

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

consists in the actual accomplishment of this endeavour (*op. cit.*, sect. 421).

"Love of our neighbor" is, in Butler, another name for benevolence, the passion for the good of others. The good of others is its object, but yet he says that in a certain way it can be satisfied by something else altogether, the very possession and exercise of the passion. This is not satisfaction in the sense that is present in the notion of a satisfaction view of happiness—the obtaining of the object of a passion or desire. If this other sort of satisfaction can be at least part of the balanced satisfaction of the passions that is happiness, then according to Butler at least some measure of well-being can come from certain concerns themselves, apart from whether they are met.[24]

These passages are inconclusive, but suggestive of some measure of agreement with Aristotle that happiness is the exercise of virtue —a character view of happiness, not a satisfaction view. If this is Butler's view, then it is not troubled by the loop we have been discussing. And whatever Butler may have believed, the loop problem appears to put pressure on satisfaction views of well-being, pressure that does not apply to the view that well-being is a matter of character, nor to other nonsatisfaction views, such as hedonist or objective-list views. As for which model of happiness is ultimately superior, this is only one among many important considerations.

DAVID M. ESTLUND

University of California/Irvine

[24] That he still speaks of this case in terms of gratification should make us cautious in our interpretation of his other satisfaction-like language. It does not necessarily indicate the presence of a satisfaction model of happiness.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

COMMENTS AND CRITICISM

A SECOND LOOK AT PORNOGRAPHY AND THE SUBORDINATION OF WOMEN

FEMINIST writer Andrea Dworkin and feminist lawyer Catherine MacKinnon recently proposed a new definition of "pornography" as "the graphic sexually explicit subordination of women" (henceforth D/M).[1] They introduced this definition as a part of an ordinance which was designed to help the city of Minneapolis control its burgeoning adult entertainment industry. The ordinance has not met with much legal success. (The Minneapolis city council barely passed it, but the Mayor vetoed it. And the United States Supreme Court summarily dismissed it.) Nonetheless, its conception of pornography is philosophically intriguing and worthy of analysis.

Melinda Vadas[2] attempts an ingenious, highly original defense of the Dworkin/MacKinnon definition. She concedes that her task will not be an easy one. To equate pornography with the graphic, sexually explicit subordination of women "should make any English speaker sit bolt upright" (489). Would it not be much more reasonable, she asks, to conceive of pornography as the graphic, sexually explicit *depiction* of the subordination of women?

On this reasonable view (as Vadas herself calls it) pornography is not an intrinsic injustice. It may be highly offensive to many, and producing it may be a bad thing to do. But we can go no further in

[1] The full statement of the definition is as follows:

1. Pornography is the graphic sexually explicit subordination of women through pictures and/or words that also includes one or more of the following: (i) women are presented dehumanized as sexual objects, things, or commodities; or (ii) women are presented as sexual objects who enjoy pain or humiliation; or (iii) women are presented as sexual objects who experience sexual pleasure in being raped; or (iv) women are presented as sexual objects tied up or cut up or mutilated or bruised or physically hurt; or (v) women are presented in postures or positions of sexual submission, servility, or display; or (vi) women's body parts—including but not limited to vaginas, breasts, or buttocks—are exhibited such that women are reduced to those parts; or (vii) women are presented as whores by nature; or (viii) women are presented being penetrated by objects or animals; or (ix) women are presented in scenarios of degradation, injury, torture, shown as filthy or inferior, bleeding, bruised or hurt in a context that makes these conditions sexual.
2. The use of men, children, or transsexuals in the place of women in (1) above is pornography for the purposes of this law.

[2] "A First Look at the Pornography/Civil Rights Ordinance: Could Pornography Be the Subordination of Women?" in this JOURNAL, LXXXIV, 9 (September 1987): 487–511.

0022-362X/90/8704/205–211
© 1990 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

our moral condemnation of it. In Vadas's words: "the reasonable view does not have conceptual room for naming the production or dissemination of pornography as an injustice, but can at most see such actions as producing a state of mere disvalue" (510). D/M, on the other hand, construes pornography as a species of sexual, and sexist, subjugation which by its very nature does injustice against women.

Vadas's attempt to defend D/M utilizes a distinction between what she calls direct predicate transfers and prepositional predicate transfers. A predicate transfers directly from an object, event, or action to a depiction of that object, event, or action if it applies both to the object, event, or action and to its depiction without equivocation or other change in meaning, and if it applies to the depiction because it applies to the action, object, or event (498). On the other hand, there are predicates which make this transfer only with the aid of a preposition like 'of' or 'about'. These predicates transfer prepositionally. Some value predicates—e.g., 'good', 'bad', 'ugly', 'beautiful'—transfer directly, others only prepositionally. Vadas's object is to show that 'subordinates' falls into the former class and that as a consequence D/M is logically defensible.

Let us look carefully at her examples of direct predicate transfer.

(1) The predicate 'is red' applies to a red rose and may also apply, in the same sense of 'is red', to a depiction of the rose. And it applies to the depiction because it applies to the rose. The picture of the rose is red because the rose itself is. Vadas contrasts 'is red' with 'is sweet-smelling'. The latter obviously transfers only prepositionally. A photograph of a rose is not itself sweet-smelling. Of course, perfume could be spilled on it, but then the predicate would not apply to the photo because it applies to the rose (499).

(2) The value predicate 'is beautiful' transfers directly from a rose in nature to a photograph of the rose. The term applies to both the rose and its picture without equivocation or change in meaning, and it applies to the picture because it applies to the rose (502).

(3) Clara mistakenly drinks the water in her finger bowl and then exclaims, before all the guests at a party, "Great soup!" This action disgraces Clara. Elwood videotapes Clara's gaffe. The predicate 'disgraces Clara' transfers directly from the action itself to its depiction on film, since Vadas's two criteria are met. Or so she asserts (502).

(4) A baseball player has the fans applauding his home run. Their applause honors him. A photograph taken of the applause also honors the player. The predicate 'honors the player' transfers directly from the action of applauding to its depiction. There is, Vadas claims, no change in the meaning of the predicate, and the photo honors the player because the action of applause honored him (503). Similarly, the military honors a soldier by awarding him a medal. A depiction of the soldier wearing his medal also honors him (503).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The most striking, and discouraging, aspect of Vadas's treatment of these cases is the prevalence of assertion over argument. Perhaps she thinks that the phenomenon of direct predicate transfer is so incontrovertible that it requires no argument. But she is wrong.

Regarding case (1), all that we are warranted to say is that the predicate 'is red' applies both to the real rose and to the rose-as-it-appears in the picture. To say that the pictured rose is red, or that the rose appearing in the picture is red, is much different from claiming that the picture itself of the rose is red. How odd it would be to hand someone a photograph of a rose and say that it, the photograph, is red. (If I took a picture of a red rose and dipped it in black paint, would we want to aver that the picture itself was red?) On the other hand, it is quite proper and intelligible to say that the picture is *of* a red rose. The difference between 'is red' and 'is sweet-smelling' lies not in the fact that the former transfers directly to photographs of roses while the latter transfers only prepositionally. Rather, the difference is that the picture can directly capture visual but not olfactory qualities.

With regard to (2), 'is beautiful' does not transfer to a picture of a rose because the pictured rose is beautiful. The aesthetic criteria pertinent to the evaluation of a photograph's beauty involve much more than the beauty of the objects photographed. They include such technical matters as lighting, angle, shadow, and background. Hence, Vadas's second condition of direct predicate transfers is not satisfied in (2).

Vadas fails to provide an argument in (3) for believing that Elwood's videotape of Clara itself disgraces her. Were Elwood to show the tape to a group of friends, he might well succeed in humiliating Clara. But then it would be the use to which he put the tape and not the tape itself that disgraces her. It is difficult to understand how a videotape that is never shown to anyone could disgrace anyone. Therefore, it is difficult to understand how 'disgrace' transfers directly from Clara's action to the film of it.

As for (4), the photograph of the applause can certainly be used as a means for honoring the player. Thus, I might pass the picture around to an admiring audience as part of my tribute to his batting power. But here it is not the photo itself but my use of it in the appropriate circumstances which does the honoring. Inanimate objects do not intelligibly function as subjects for verbs like 'disgrace', and 'honor'. It makes no logical sense to claim that films, pictures, statues, books, etc., honor or disgrace people. People honor or disgrace people, sometimes by using or taking pictures, sometimes by building statues, sometimes by writing books. Likewise, the picture of the soldier wearing his medal cannot by itself honor him, though once again it can be used in the act of honoring him.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Examples (1)–(4), then, offer no support for the thesis that some predicates transfer directly from actions, objects, or events to depictions of those actions, objects, or events. Vadas's argument in support of D/M depends crucially on the claim that 'subordinates' is a value predicate that transfers directly to pornographic depictions. Unfortunately, she does not furnish an argument on behalf of this claim. Instead she is content to point out that 'subordinates', in the sense of "placing a person into a class of intrinsic moral inferiors," unquestionably constitutes a value predicate and as such can function in direct predicate transfers. Accordingly, pornography can subordinate women (508–9). D/M is not logically incongruous.

The first point to emphasize here is that we have yet to be given any reason for supposing that pornography in fact does contribute to the creation of women as a subordinated people. Consequently we have as yet no reason to accept the Dworkin/MacKinnon account of the wrongness of producing pornography, namely "it is the sexual subordination of women that *is* pornography which grounds and explains the faulty aspect of producing or disseminating it" (511—the emphasis is mine).

But does it even make any sense to say that pornographic depictions subordinate women? Of course, it is easy to think of cases where pornography might be used to facilitate such subordination. For example, a man gets excited watching a violent rape depicted graphically in an adult film and then proceeds to subdue forcibly his wife. But it is not easy to understand how books, magazines, or films can by themselves "place a person into the class of intrinsic moral inferiors." Subordination is, after all, an action or a practice engaged in by human beings and directed against other beings. In other words, the logic of 'subordinates' requires that it have some human action or actions as a subject.[3]

It also precludes characterization by the adjectives 'graphic' and 'sexually explicit'. Depictions of human behavior are often quite meaningfully called graphic and/or sexually explicit. Pornographic films and literature, by virtue of their representative content, are aptly so described. The acts they depict, however, while sexual, are not themselves sexually explicit or graphic. These words apply only to the way in which these acts are described or portrayed. Thus, we watch an adult film that shows everything in vivid color, or we read an adult novel that is replete with explicit language.

[3] Dworkin and MacKinnon at times recognize this fact. During the Minneapolis hearings, for example, they stated that pornography is a systematic *practice* of exploitation and subordination based on sex which differentially harms women. See MacKinnon, *Feminism Unmodified* (Cambridge: Harvard, 1987), p. 212. See also Paul Brest and Ann Vandenberg, "Politics, Feminism, and the Constitution: The Anti-Pornography Movement in Minneapolis," *The Stanford Law Review*, XXXIX (February 1987), p. 619.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Vadas's effort to defend D/M commits several times over what Gilbert Ryle[4] famously and felicitously dubbed a category mistake. These are mistakes which arise from the inability to use correctly certain words in our vocabulary. Ryle writes:

> the theoretically interesting category mistakes are those made by people who are perfectly competent to apply concepts, at least in the situations with which they are familiar, but are still liable in their abstract thinking to allocate these concepts to logical types to which they do not belong (*ibid.*, p. 17).

Vadas's assertions about photographs honoring ballplayers and videotapes disgracing guests are logically odd in just the way that Ryle describes. Honor and disgrace do not belong, logically, to pictures or tapes. Similarly, books, magazines, and films are not the kinds of things that can meaningfully be said to subordinate.

Rejecting D/M, must we then agree with Vadas's claim that, on the reasonable view, pornography comprises just a "mere disvalue"? Let us for the sake of argument accept a narrow conceptualization of pornography as consisting of verbal or pictorial material that depicts the coerced or violent subordination of women. (This is narrow because it rules out by definition pornography in which men are shown as victimized as well as so called erotic pornography in which explicit sexual relations are depicted as expressions of love and caring between consenting partners.) What makes this species of pornography morally bad?

Some feminists claim that it is condemnable as anti-female propaganda.[5] But there is, to my knowledge, no research data indicating that exposure to pornography produces a contemptuous attitude toward women. Nor is there good reason to believe that the producers of pornography are determined to propagate such an attitude in their customers as opposed to exploiting an already existing misogynism for commercial gain. Explicit depictions of women being raped, beaten, tortured, hung from ceilings, and in other ways brutally subjugated are not likely to acquire entertainment or educational value in persons who like or respect women. On the contrary, these persons will be repulsed by such pornography.

Does exposure to pornography beget violence against women? Empirical studies of this question run up against the intractable problem of distinguishing causal claims from claims that watching or reading pornography simply reinforces preexisting negative attitudes toward women. Again, the idea that a man who harbors no

[4] *The Concept of Mind* (New York: Barnes & Noble, 1949).

[5] See, for example, Susan Brownmiller's *Against Our Will: Men, Women, and Rape*, excerpted in Laura Lederer, ed., *Take Back the Night* (New York: Morrow, 1980), p. 33.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

disdain for women will become disposed to harm them after reading or viewing pornography invites the skeptical response: Why would utter disgust and disdain for pornographers not be his natural response to this kind of material? The causal argument also has to explain the fact that in Denmark, Sweden, and the Netherlands anti-women violence is relatively uncommon despite the widespread availability of pornography, while in Ireland and South Africa such violence is relatively common even though pornography is unavailable.[6]

So the reasonable view will abjure the suggestion that the major disvalue of pornography consists of either sexist propaganda or the incitement of violence against women. Instead it construes the main function of pornography, in the narrow sense under discussion here, to be misogynist pandering. And it identifies this function as the principal source of pornography's disvalue.

'To pander' means "to minister to the baser passions of another." The baser passion to which pornography ministers is not an interest in sex per se. Mature sexual behavior free of duplicity, coercion, and violence is not morally objectionable. Violent subordination through sex is morally objectionable, though, and it is precisely an interest in this kind of debasement to which pornography panders.

The use of 'subordination' in this context should not cause a conceptual difficulty, but Vadas's conceptualization of it does. She avers that subordination involves placing a person or group of persons socially in the class of those whose intrinsic moral worth is not of the first rank (506). Unfortunately, the idea of intrinsic moral worth Vadas deploys here needs clarification, particularly in light of her view that bearers of this worth can be ranked in some kind of hierarchy. Ordinarily, the appeal to intrinsic worth serves to rule out any such ranking.

A more common and perspicuous account of subordination equates it with the practice of regarding a person as inferior to others in some respect(s) and as a consequence placing him/her in a lower position, group, or class. There is nothing intrinsically wrong with this practice. Among its justifying purposes are greater technical knowledge, wider experience, and proven maturity of judgment. Still, many forms of subordination are morally unjust, and prominent among these is the arbitrary, brutal exercise of power by men over women depicted and endorsed in pornography.

The most appropriate and most forceful characterization of materials which describe or portray sexist subordination is that they are morally evil. 'Moral evil' is a strong term of moral condemnation. Vadas claims that the reasonable view of pornography can at most

[6] Joel Feinberg discusses this point in his *Offense to Others* (New York: Oxford, 1986), p. 149.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

criticize it as a "mere disvalue" (510). But moral evil is much more than a mere disvalue.

According to the *Oxford English Dictionary,* 'evil' means first and foremost "morally depraved." 'To deprave', in turn, means "to debase, vilify, or disparage." And 'to debase' means "to lower in dignity, to depreciate, to bring into contempt." So there is an inextricable connection between the ideas of evil and the contemptuous treatment of persons. John Rawls[7] is right on the conceptual and moral mark when he writes: "What moves the evil man is the love of injustice. He delights in the impotence and humiliation of those subject to him, and he relishes being recognized by them as the wilful author of their degradation" (*ibid.*, p. 439).

The most insidious evil actions are those which target a specific group for systematic devaluation and humiliation: Nazi terrorism against Jews, white supremacists' propaganda against blacks, social discrimination aimed at the old and sick, misogynists' abuse of women. But moral evil need not involve the actual exercise of unjust power. It is also found in the practice of profiting from such exercise—e.g., slave ownership. 'Moral evil' also properly applies to the activity of producing verbal or pictorial representations of evil acts for the sake of others' amusement. Thus, it would be morally evil to make a film celebrating Nazi atrocities for the viewing pleasure of anti-Semites. Finally, the pleasure taken in depictions of human debasement is most appropriately described as evil, and is integral to the evil person's character (Rawls's point).

The application of 'moral evil' to pornography is clear enough. Materials that depict in a favorable light the arbitrary debasement of women are intrinsically contemptuous of women and as such are morally evil. Likewise, the pleasure a person takes in reading or watching pornography is evil. Hence, pandering to this pleasure for the sake of commercial gain is an evil practice. So the business of pornography, which I equated earlier with a form of ignoble pandering, is on the reasonable view condemnable in very strong terms. It thrives by exploiting the profoundly pernicious enjoyment too many men find in pornographic images of demeaning subordination. Opponents of pornography should focus on this important point instead of proferring new but philosophically indefensible definitions of 'pornography'.

W. A. PARENT
Santa Clara University

[7] *A Theory of Justice* (Cambridge: Harvard, 1971).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

BOOK REVIEWS

The Intentional Stance. DANIEL DENNETT. Cambridge: MIT Press, 1987. xi, 388 p. Cloth $27.50, paper $11.95.

Readers of Daniel Dennett's earlier works will know what virtues to expect in this welcome book. They will know that Dennett writes with more wit and flair and zest than anyone else on the current philosophical scene, that his work is keenly intelligent and exceptionally insightful, that he is provocative (and occasionally provoking), that he tramples disciplinary boundaries, thereby illuminating philosophical issues by bringing to bear on them an impressive knowledge of other subjects (cognitive science, ethology, evolutionary biology, etc.), that he makes masterful use of examples (or "intuition pumps," as he likes to call them), and that he is death on sacred cows (though perhaps a bit given to tilting at windmills).

A number of the essays in the volume, in particular "True Believers" (which Dennett now regards as the "flagship expression" of his position), "Three Kinds of Intentional Psychology," and "Beyond Belief," were well-known and widely discussed prior to their republication here. The six previously published essays are accompanied by "reflections" in which Dennett responds to criticisms and tries to clarify and integrate their central themes. There are two additional papers, published here for the first time, which develop his ideas in new directions (most notably, evolutionary biology), and a concluding chapter, "Midterm Examination: Compare and Contrast," which gives us a Dennett's eye view of the recent history of theories of intentionality, beginning with the work of Wilfrid Sellars and W. V. Quine and culminating in the Dennettian conception of intentional attribution, with Jerry Fodor the odd man out.

The basic outlines of Dennett's position on intentional attribution are well-known. We take the "intentional stance" toward something when we regard it as being rational, as having the beliefs and desires it "ought" to have, and as being disposed to behave accordingly. Something is an intentional system to the extent that taking the intentional stance toward it is a useful, or essential, way of explaining and predicting its behavior. The beliefs, etc., such a system has are those which would be attributed to it in a successful employment of the intentional stance. There is no sharp line between things that are intentional systems and things that are not, or between "real believers" and things like thermostats.

0022-362X/90/8704/212–216 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Dennett now says that it was a "tactical mistake" for him to accept the characterization of this position as instrumentalist, but he makes little effort to make it seem less so. He insists that his position is "realist" (small 'r') about belief, but not "Realist" (large 'R') in the sense Fodor's is. Part of what he rejects in rejecting "Realism" is the language-of-thought thesis (although that rejection is less firm in some places than in others). But what most of all makes his position seem not a realist one (large 'r' or small), and instrumentalist, is his denial that beliefs, etc., play, or at least need be thought of as playing, any role in the *causation* of behavior. He distinguishes (following Hans Reichenbach) between *illata,* which are genuine theoretical entities (e.g., electrons), and *abstracta,* which are "calculation bound entities or logical constructs" (e.g., centers of gravity, and the equator), and thinks that beliefs and the like are best placed among the latter. Apparently he thinks that folk psychology tries to have it both ways, and in "Three Kinds of Intentional Psychology" he advocates two successor theories: *intentional systems theory,* in which beliefs, etc., are consistently regarded as *abstracta,* and *subpersonal cognitive psychology,* which invokes states having intentional content that are genuine *illata,* and which do have causal efficacy, but which cannot be identified with beliefs or other intentional states of folk psychology. (Despite its grand name, intentional systems theory seems to have no content beyond the platitudes of folk psychology —perhaps one could convey its contents by citing these platitudes and then adding 'understand these instrumentally'.)

Dennett says little in *The Intentional Stance* about the distinction between *illata* and *abstracta,* but he makes some revealing remarks about this in a reply[1] to Fred Dretske.

> There is always more to learn about things we are realists about: electricity or electrons, DNA molecules or chromosomes; but there really isn't anything more to learn about centers of gravity or Mendelian genes. . . . They are *abstracta,* and what there is to learn about them is just whatever follows from their combinatorial roles in our theories (*ibid.,* p. 537).

It would follow from this, I think, that functional states, conceived under functional descriptions that are definitive of them, are *abstracta;* we can of course learn empirically various ways in which such a state can be realized or implemented (as we might also do with Mendelian genes), but presumably that will not count as learning

[1] Dennett's response to peer commentary on *The Intentional Stance,* in *Behavioral and Brain Sciences,* 1, 3 (September 1988): 535–544.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

more about the state itself, in the sense Dennett has in mind. If this is the status being assigned to beliefs, well and good. But being *abstracta* in this sense does not preclude having causal efficacy—not if functional states are defined by causal roles. Nor can I see why such a view of beliefs need be incompatible with robust realism about them (to be a realist about chairs one does not have to think that 'chair' has a semantics like that of natural kind terms, as understood by Saul Kripke and Hilary Putnam, and that there is always something more to learn about chairs, qua chairs).

It seems plain that any spelling out of the principles we tacitly take for granted in employing the intentional stance (e.g., any spelling out of the consequences of the rationality assumption) would involve asserting all sorts of counterfactual dependence of these states on one another and on environmental conditions, and of overt behavior on combinations of these states. If this counterfactual dependence does not rest on causal dependence, one would like to know why. Dennett clearly thinks that, if we want a scientifically revealing account of why the intentional stance works in particular cases, we must go to subpersonal cognitive psychology. And he may think that, if we see the *illata* invoked there as causes of behavior, we are barred from viewing the states of folk psychology as causes of what is in some sense the same behavior. But that would be a mistake. The *illata* of subpersonal folk psychology will presumably be realized in neurophysiological states, events, and processes. Dennett will certainly want to allow that there is some description of the behavior under which it can appropriately be said to be caused by these neurophysiological states, etc., and that it is consistent with this that there is some description of it under which it can appropriately be said to be caused by the *illata* of subpersonal cognitive psychology. So why should there not also be a description of it under which it can appropriately be said to be caused by folk psychological states, e.g., beliefs and desires, which are realized in the subpersonal *illata,* which in turn are realized in the neurophysiological states (all of this realization being, presumably, of a highly holistic nature)?

Dennett's chief argument against Realism about beliefs appears to be that they lack well-defined identity conditions—that there is no secure basis for a sharp distinction between beliefs that are "explicitly" stored and those stored only "implicitly." But this seems mainly a problem for language-of-thought accounts of such states; and, in any case, it is far from clear why lack of sharp identity conditions should be thought a reason for anti-Realism of a kind that denies causal efficacy.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Another apparent source of Dennett's anti-Realism about folk psychological intentional states is the role the "idealizing" assumption of rationality is supposed to play in the intentional stance. He repeatedly lays stress on this, saying at one point that it is the "myth of our rational agenthood that structures and organizes our attributions of belief and desire to others and that regulates our own deliberations and investigations" (52). Yet the import of this is left obscure, for he denies that he regards beliefs, etc., as fictions. In any case, the use of 'myth' here seems unwarranted; there seems no good reason to think that our employment of the intentional stance requires attributing more rationality to people than they typically have.

Dennett's instrumentalism and anti-Realism are only one strand in his view about intentionality, and he has many illuminating things to say about the centrality to our conceptual scheme, and the indispensability, of the intentional stance. A favorite passage of mine is one, in "True Believers," in which he discusses the example (taken from Robert Nozick) of Martian superscientists who predict human behavior from the "physical stance," and compares their explanation of a typical bit of human behavior with our "intentional stance" explanation of it. This could hardly be improved on as an expose of a certain sort of reductionist (or eliminativist) illusion.

"Evolution, Error, and Intentionality," one of the papers not previously published, is perhaps the most provocative essay in the book. Here Dennett develops a very illuminating analogy between intentional explanation and evolutionary explanation—which he does not quite succeed in ruining by linking it with what seemed to me an ill-conceived assault on the view that human intentionality is "original" and "intrinsic" (the word 'intrinsic' seems to be a red flag for Dennett), and by going beyond the claim of mere analogy to maintain that we are designed by mother nature, a.k.a., the process of natural selection, in just the sense (he seems to be saying) in which artifacts are designed by us.

The "bad guys" (his term) of this essay are those who believe that there is a distinction between "original," or "intrinsic," intentionality, and "derived" intentionality, the former belonging to us and the latter belonging to artifacts. These include John Searle, Fodor, Dretske, Tyler Burge, and Kripke, along with Roderick Chisholm, Thomas Nagel, and Karl Popper and John Eccles. The "good guys," who "find this intuition dubious if not downright dismissible," include, besides Dennett himself, Patricia and Paul Churchland, Donald Davidson, John Haugeland, Ruth Millikan, Richard Rorty, and Robert Stalnaker, along with Quine and Sellars and "almost every-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

one in AI." These lists give us some strange bedfellows, and I am not convinced that there is a single (relevant) issue that unites the people on each of them and divides those on the one from those on the other. And Dennett's discussion seems to me to conflate at least four distinct questions. (1) Is there a viable distinction between intentionality that is derived, in the sense that something has it in virtue of the purposes, intentions, etc., of its designers or users (e.g., that of the speedometer reading on my car), and intentionality that is original in the sense of not being so derived, such that human intentionality is of the latter sort? (2) Do evolutionary considerations play a role in determining the intentional content of the states of humans and other organisms, analogous to that played by human purposes in determining the intentional content of states of artifacts? (3) Can artifacts have original intentionality? (4) Can it be indeterminate what the intentional content of a state of a human (or other organism) is? Dennett appears to assume, without any warrant that I can see, that anyone who answers *yes* to (1) must answer *no* to all of (2)–(4). For the record, I think that it is obvious beyond argument that the answer to (1) is *yes* (so I am a "bad guy"), and think that it is arguable that some or all of the others should be answered *yes* as well. I have no fixed opinion about the answer to (2); but I think that those who answer it *yes* should not allow themselves to be stuck with the view that in so doing they are taking the intentional stance toward nature as a whole. Sometimes an analogy is no more than an analogy.

Essays I have not had space to discuss are "Making Sense of Ourselves" (a reply to Stephen Stich in which Dennett clarifies his position on the rationality assumption), "Beyond Belief" (where he has his say on the issues raised by Twin Earth type examples), "Styles of Mental Representation," "Intentional Systems in Cognitive Ethology: The 'Panglossian Paradigm' Defended" (where he takes on Steven Jay Gould and Richard Lewontin on the issue of "adaptationism," and also reports first-hand on the vervets), and "Fast Thinking" (an attack on Searle's Chinese Room argument). All will raise hackles, and all are warmly recommended.

SYDNEY SHOEMAKER

Cornell University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NEW BOOKS

ADAMS, IAN: *The Logic of Political Belief: A Philosophical Analysis of Ideology.* Savage, MD: Barnes & Noble, 1989. xiii, 155 p. Cloth $39.75.

ADLER, MORTIMER J.: *Intellect: Mind over Matter.* New York: Macmillan, 1990. 184 p. Cloth $16.95.

ALSTON, WILLIAM P.: *Epistemic Justification: Essays in the Theory of Knowledge.* Ithaca: Cornell, 1989. xi, 354 p. Cloth $44.95, paper $14.95.

ANDERSON, DAVID: *Artificial Intelligence and Intelligent Systems: The Implications.* New York: Wiley, 1989. 178 p. Cloth $49.95.

ARMSTRONG, D. M.: *A Combinatorial Theory of Possibility.* New York: Cambridge, 1989. xiii, 156 p. Cloth $34.50, paper $11.95.

ARRINGTON, ROBERT L.: *Rationalism, Realism, and Relativism: Perspectives in Contemporary Moral Epistemology.* Ithaca: Cornell, 1989. 321 p. Cloth $36.50, paper $13.95.

AUSTIN, DAVID F.: *What's the Meaning of "This"?: A Puzzle about Demonstrative Belief.* Ithaca: Cornell, 1990. xvii, 172 p. Cloth $22.95.

AUSTIN, NORMAN: *Meaning and Being in Myth.* University Park: Pennsylvania State UP, 1990. xiii, 239 p. Cloth $28.50.

BACH, EMMON: *Informal Lectures on Formal Semantics.* Albany: SUNY, 1989. 150 p. Cloth $29.50, paper $9.95.

BALABAN, ODED: *Subject and Consciousness: A Philosophical Inquiry into Self-Consciousness.* Savage, MD: Rowman, 1990. xi, 221 p. Cloth $32.50.

BARTLETT, RANDALL: *Economics and Power: An Inquiry into Human Relations and Markets.* New York: Cambridge, 1989. xi, 209 p. Cloth $32.50.

BELL, LINDA A.: *Sartre's Ethics of Authenticity.* University: Alabama UP, 1989. 223 p. Cloth $23.95.

BENJAMIN, ANDREW: *Translation and the Nature of Philosophy: A New Theory of Words.* New York: Routledge, 1989. v, 193 p. Cloth $49.95, paper $14.95.

BENJAMIN, MARTIN: *Splitting the Difference: Compromise and Integrity in Ethics and Politics.* Lawrence: Kansas UP, 1990. x, 195 p. Cloth $22.50, paper $12.95.

BENNET, BRUCE M., DONALD D. HOFFMAN, and CHETAN PRAKASH: *Observer Mechanics: A Formal Theory of Perception.* New York: Academic, 1989. xii, 285 p. Cloth $29.95.

BLACK, MAX: *Perplexities: Rational Choice, the Prisoner's Dilemma, Metaphor, Poetic Ambiguity, and Other Puzzles.* Ithaca: Cornell, 1990. xi, 201 p. Cloth $25.95.

BLUM, LAWRENCE A., and VICTORIA J. SEIDLER: *A Truer Liberty: Simone Weil and Marxism.* New York: Routledge, 1989. xxii, 346 p. Paper $17.95.

BOGUE, RONALD: *Deleuze and Guattari.* New York: Routledge, 1989. xiii, 196 p. Cloth $42.50, paper $13.95.

BOOLOS, GEORGE S., and RICHARD C. JEFFERY: *Computability and Logic.* New York: Cambridge, 1989, 3rd ed. xi, 304 p. Cloth $49.50, paper $17.95.

BOTWINICK, ARYEH: *Skepticism and Political Participation.* Philadelphia: Temple, 1990. xiii, 264 p. Cloth $34.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

BOURGEOIS, PATRICK L., and FRANK SCHALOW: *Traces of Understanding: A Profile of Heidegger's and Ricoeur's Hermeneutics.* Atlanta: Rodopi, 1990. vi, 186 p.

BOWEN-MOORE, PATRICIA: *Hannah Arendt's Philosophy of Natality.* New York: St. Martin's, 1989. x, 190 p. Cloth $24.95.

BROADIE, ALEXANDER: *Notion and Object: Aspects of Late Medieval Epistemology.* New York: Oxford, 1989. 197 p. Cloth $45.00.

BROMWICH, DAVID: *A Choice of Inheritance: Self and Community from Edmund Burke to Robert Frost.* Cambridge: Harvard, 1989. Cloth $29.50.

BROWN, JUDITH: *Gandhi: Prisoner of Hope.* New Haven: Yale, 1990. xiv, 440 p. Cloth $29.95.

BRUNS, GERALD L.: *Heidegger's Estrangements: Language, Truth, and Poetry in the Later Writings.* New Haven: Yale, 1989. xxx, 233 p. Cloth $26.50.

BUCK MOSS, SUSAN: *The Dialectics of Seeing.* Cambridge: MIT, 1989. xii, 493 p. Cloth $29.95.

BUDD, MALCOLM: *Wittgenstein's Philosophy of Psychology.* New York: Routledge, 1989. xiv, 186 p. Cloth $39.95.

BUSCH, THOMAS W.: *The Power of Consciousness and the Force of Circumstances in Sartre's Philosophy.* Bloomington: Indiana UP, 1990. xiv, 112 p. Cloth $19.95.

BUTLER, JUDITH: *Gender Trouble: Feminism and the Subversion of Identity.* New York: Routledge, 1989. xiv, 172 p. Paper $12.95.

CALLAHAN, DANIEL: *What Kind of Life: The Limits of Medical Progress.* New York: Simon & Schuster, 1990. 318 p. Cloth $19.95.

CARR, THOMAS M.: *Descartes and the Resilience of Rhetoric: Varieties of Cartesian Rhetorical Theory.* Carbondale: Southern Illinois UP, 1989. xi, 218 p. Cloth $22.95.

CARROLL, NOEL: *The Philosophy of Horror or Paradoxes of the Heart.* New York: Routledge, 1990. xi, 256 p. Cloth $45.00, paper $14.95.

CARTER, WILLIAM R.: *The Elements of Metaphysics.* Philadelphia: Temple, 1989. x, 195 p. Cloth $29.95.

CLARK, STEPHEN R. L.: *Civil Peace and Sacred Order: Limits and Renewels I.* New York: Oxford, 1989. x, 198 p. Cloth $32.50.

CLARKE, DESMOND M.: *Occult Powers and Hypotheses: Cartesian Natural Philosophy under Louis XIV.* New York: Oxford, 1989. x, 265 p. Cloth $64.00.

CLULEE, NICHOLAS H.: *John Dee's Natural Philosophy: Between Science and Religion.* New York: Routledge, 1989. xiv, 347 p. Cloth $55.00, paper $17.95.

COHEN, G. A.: *History, Labour and Freedom: Themes from Marx.* New York: Oxford, 1989. xvii, 317 p. Cloth $65.00, paper $24.95.

CONNOR, STEVEN: *Postmodernist Culture: An Introduction to Theories of the Contemporary.* New York: Blackwell, 1989. ix, 274 p. Cloth $39.95, paper $14.95.

COPPER-WIELE, JONATHAN K.: *The Totalizing Act: Key to Husserl's Early Philosophy.* Boston: Kluwer, 1989. x, 150 p. Cloth $59.00.

CUMMINS, ROBERT: *Meaning and Mental Representation.* Cambridge: MIT, 1989. viii, 180 p. Cloth $20.00.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

DALLMAYR, FRED: *Margins of Political Discourse.* Albany: SUNY, 1989. xvii, 271 p. Cloth $54.50, paper $17.95.

DENT, N. J. H.: *Rousseau: An Introduction to His Psychological, Social and Political Theory.* New York: Blackwell, 1989. xiii, 258 p. Cloth $55.00.

DEWART, LESLIE: *Evolution and Consciousness.* Toronto: University Press, 1989. xii, 399 p. Cloth $50.00.

DINWIDDY, JOHN: *Bentham.* New York: Oxford, 1989. viii, 132 p. Paper $7.95.

DORE, CLEMENT: *God, Suffering and Solipsism.* New York: St. Martin's, 1989. x, 120 p. Cloth $35.00.

EAMES, ELIZABETH RAMSDEN: *Bertrand Russell's Dialogue with His Contemporaries.* Carbondale: Southern Illinois UP, 1989. xvi, 287 p. Cloth $34.95.

EARMAN, JOHN: *World Enough and Space-Time: Absolute versus Relational Theories of Space and Time.* Cambridge: MIT, 1990. xiv, 223 p. Cloth $25.00.

ECCLES, JOHN C.: *Evolution of the Brain: Creation of the Self.* New York: Routledge, 1989. xv, 282 p. Cloth $25.00.

EDEL, ABRAHAM: *Interpreting Education.* Buffalo: Prometheus, 1989. x, 340 p. Paper $18.95.

ELRIDGE, RICHARD: *On Moral Personhood: Philosophy, Literature, Criticism, and Self-Understanding.* Chicago: University Press, 1989. xiv, 210 p. Cloth $27.50.

ERSKINE, ANDREW: *The Hellenistic Stoa: Political Thought and Action.* Ithaca: Cornell, 1990. xii, 233 p. Cloth $37.50.

FALCK, COLIN: *Myth, Truth and Literature: Towards a True Post-Modernism.* New York: Cambridge, 1989. xv, 173 p. Cloth $34.50.

FARARO, THOMAS J.: *The Meaning of General Theoretical Sociology: Tradition & Formalization.* New York: Cambridge, 1989. xi, 387 p. Cloth $42.50.

FARRAR, CYNTHIA: *The Origins of Democratic Thinking: The Invention of Politics in Classical Athens.* New York: Cambridge, 1989. x, 301 p. Cloth $37.50, paper $14.95.

FISK, MILTON: *The State and Justice: An Essay in Political Theory.* New York: Cambridge, 1989. x, 391 p. Cloth $49.50, paper $16.50.

FLAX, JANE: *Thinking Fragments: Psychoanalysis, Feminism, Postmodernism in the Contemporary West.* Berkeley: California UP, 1990. ix, 277 p. Cloth $25.00.

FORGUSON, LYND: *Common Sense.* New York: Routledge, 1989. vi, 193 p. Cloth $35.00.

FORMAN, FRANK: *The Metaphysics of Liberty.* Boston: Kluwer, 1989. xi, 202 p. Cloth $67.00.

FOX, RICHARD G.: *Gandhian Utopia: Experiments with Culture.* Boston: Beacon, 1989. x, 330 p. Cloth $27.50.

FRASER, NANCY: *Unruly Practices: Power, Discourse, and Gender in Contemporary Social Theory.* Minneapolis: Minnesota UP, 1989. viii, 201 p. Cloth $35.00, Paper $14.95.

GILL, MARY LOUISE: *Aristotle on Substance: The Paradox of Unity.* Princeton: University Press, 1990. xi, 284 p. Cloth $29.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

GILLETT, GRANT: *Reasonable Care.* New York: St. Martin's, 1989. xi, 147 p. Cloth $29.95.

GODLOVE, TERRY F.: *Religion, Interpretation, and Diversity of Belief: The Framework Model from Kant to Durkheim to Davidson.* New York: Cambridge, 1989. xii, 207 p. Cloth $34.50.

GRICE, PAUL: *Studies in the Way of Words.* Cambridge: Harvard, 1989. vi, 394 p. Cloth $32.50.

HANCOCK, RALPH C.: *Calvin and the Foundations of Modern Politics.* Ithaca: Cornell, 1989. xvii, 221 p. Cloth $26.50.

HELD, VIRGINIA: *Rights and Goods: Justifying Social Action.* Chicago: University Press, 1989. vi, 327 p. Cloth $14.95.

HELLMAN, GEOFFREY: *Mathematics without Numbers.* New York: Oxford, 1989. xi, 154 p. Cloth $35.00.

HERBERT, GARY B.: *Thomas Hobbes: The Unity of Scientific & Moral Wisdom.* Vancouver: British Columbia UP, 1989. xiv, 201 p. Cloth $29.95, paper $15.95.

HICKMAN, LARRY A.: *John Dewey's Pragmatic Technology.* Bloomington: Indiana UP, 1990. xvii, 234 p. Cloth $29.95.

HOWSON, COLIN, and PETER URBACH: *Scientific Reasoning: The Bayesian Approach.* La Salle: Open Court, 1989. xii, 312 p. Cloth $34.95, paper $16.95.

HUGHES, CHRISTOPHER: *On a Complex Theory of a Simple God: An Investigation in Aquinas' Philosophical Theology.* Ithaca: Cornell, 1989. xi, 281 p. Cloth $34.95.

HUMPHREYS, PAUL: *The Chances of Explanation: Causal Explanation in the Social, Medical, and Physical Sciences.* Princeton: University Press, 1989. x, 170 p. Cloth $29.50.

HURLEY, S. L.: *Natural Reasons: Personality and Polity.* New York: Oxford, 1989. xii, 462 p. Cloth $45.00.

JARZOMBEK, MARK: *On Leon Baptista Alberti: His Literary and Aesthetic Theories.* Cambridge: MIT, 1989. xvi, 262 p. Cloth $19.95.

JEWELL, P. D.: *By What Authority: Anarchism, the State and the Individual.* North Adelaide: Amada, 1989. 235 p. Paper $12.00.

JOHNSON, PETER: *Politics, Innocence, and the Limits of Goodness.* New York: Routledge, 1989. 283 p. Cloth $35.00.

JOHNSTON, PAUL: *Wittgenstein and Moral Philosophy.* New York: Routledge, 1989. x, 244 p. Cloth $37.50.

JONES, HOWARD: *The Epicurean Tradition.* New York: Routledge, 1989. vii, 276 p. Cloth $39.95.

KAGAN, SHELLY: *The Limits of Morality.* New York: Oxford, 1989. xiii, 415 p. Cloth $39.95.

KELLNER, DOUGLAS: *Jean Baudrillard: From Marxism to Postmodernism and Beyond.* Stanford: University Press, 1989. ix, 246 p. Cloth $35.00, paper $11.95.

KERSZBERG, PIERRE: *The Invented Universe: The Einstein-De Sitter Controversy and the Rise of Relativistic Cosmology.* New York: Oxford, 1989. ix, 403 p. Cloth $65.00.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

KILNER, JOHN F.: *Who Lives? Who Dies?: Ethical Criteria in Patient Selection.* New Haven: Yale, 1990. xiv, 359 p. Cloth $29.95.

KIVY, PETER: *Music Alone: Philosophical Reflections on the Purely Musical Experience.* Ithaca: Cornell, 1990. xii, 226 p. Cloth $27.50; *Sound Sentiment: An Essay on the Musical Emotions.* Philadelphia: Temple, 1989. xvi, 286 p. Cloth $39.00, paper $19.95.

KNIGHT, DAVID: *A Companion to the Physical Sciences.* New York: Routledge, 1989. 177 p. Cloth $25.00.

KOHAK, ERAZIM: *Jan Patocka: Philosophy and Selected Writings.* Chicago: University Press, 1989. xiv, 386 p. Cloth $55.00, paper $21.95.

KURTZ, PAUL: *Europraxophy: Living without Religion.* Buffalo: Prometheus, 1989. 159 p. Cloth $15.95.

KYMLICKA, WILL: *Liberalism, Community, and Culture.* New York: Oxford, 1989. 280 p. Cloth $38.00.

LAKOFF, GEORGE, and MARK TURNER: *More Than Cool Reason: A Field Guide to Poetic Metaphor.* Chicago: University Press, 1989. xii, 230 p. Cloth $29.95, paper $11.95.

LEE, KEEKOK: *Social Philosophy and Ecological Scarcity.* New York: Routledge, 1989. xiv, 425 p. Cloth $55.00.

LEHRER, KEITH: *Thomas Reid.* New York: Routledge, 1989. xii, 311 p. Cloth $49.95.

LESLIE, JOHN: *Universes.* New York: Routledge, 1989. vii, 228 p. Cloth $25.00.

LEVY, ZE'EV: *Baruch or Benedict: On Some Jewish Aspects of Spinoza's Philosophy.* New York: Lang, 1990. xi, 224 p. Cloth $39.50.

LIDDON, SIM C.: *The Dual Brain, Religion, and the Unconscious.* Buffalo: Prometheus, 1989. 265 p. Cloth $21.95.

LISZKA, JAMES JACOB: *The Semiotic of Myth: A Critical Study of the Symbol.* Bloomington: Indiana UP, 1990. xi, 243 p. Cloth $29.95.

LUCAS, GEORGE R., JR.: *The Rehabilitation of Whitehead: An Analytic and Historical Assessment of Process Philosophy.* Albany: SUNY, 1989. xiv, 261 p. Cloth $44.50, paper 14.95.

LUCAS, J. R.: *The Future.* New York: Blackwell, 1989. x, 245 p. Cloth $29.95.

MACDONALD, CYNTHIA: *Mind-Body Identity Theories.* New York: Routledge, 1989. xiii, 255 p. Cloth $45.00.

MACHAN, TIBOR R.: *Liberty and Culture: Essays on the Idea of a Free Society.* Buffalo: Prometheus, 1989. 288 p. Cloth $19.95.

MACKENDRICK, PAUL: *The Philosophical Books of Cicero.* New York: St. Martin's, 1989. ix, 492 p. Cloth $49.95.

MAILLOUX, STEVEN: *Rhetorical Power.* Ithaca: Cornell, 1989. xiv, 189 p. Cloth $24.95, paper $8.95.

MALONEY, J. CHRISTOPHER: *The Mundane Matter of the Mental Language.* New York: Cambridge, 1989. xxvii, 274 p. Cloth $39.50.

MARQUAND, ODO: *Farewell to Matters of Principle.* New York: Oxford, 1989. 147 p. Cloth $27.50.

MARTIN, MICHEL: *Atheism: A Philosophical Justification.* Philadelphia: Temple, 1990. xiii, 541 p. Cloth $39.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

MARTIN, NORMAN M.: *Systems of Logic.* New York: Cambridge, 1989. xi, 318 p. Cloth $44.50, paper $15.95.

MASON, JEFF: *Philosophical Rhetoric: The Function of Indirection to Philosophical Writing.* New York: Routledge, 1989. xiii, 170 p. Cloth $45.00, paper $13.95.

MAZLISH, BRUCE: *A New Science: The Breakdown of Connections and the Birth of Sociology.* New York: Oxford, 1989. xiv, 333 p. Cloth $35.00.

McCARTHY, MICHAEL H.: *The Crisis of Philosophy.* Albany: SUNY, 1989. xxi, 283 p. Cloth $59.50, paper $19.95.

McCULLOCH, GREGORY: *The Name of the Game: Introducing Logic, Language, and Mind.* New York: Oxford, 1989. xvi, 320 p. Cloth $55.00, paper $17.95.

McGRANE, BERNARD: *Beyond Anthropology: Society and the Other.* New York: Columbia, 1989. x, 150 p. Cloth $28.00.

MEYER, LEONARD B.: *Style and Music: Theory, History, and Ideology.* Philadelphia: Pennsylvania UP, 1989. xi, 376 p. Cloth $39.95.

MOORE, KATHLEEN DEAN: *Pardons: Justice, Mercy, and the Public Interest.* New York: Oxford, 1989. x, 271 p. Cloth $22.95.

MOSER, PAUL K.: *Knowledge and Evidence.* New York: Cambridge, 1989. xii, 285 p. Cloth $39.50.

MULLETT, MICHAEL: *Calvin.* New York: Routledge, 1989. ix, 68 p. Paper $7.95.

MUNITZ, MILTON K.: *The Question of Reality.* Princeton: University Press, 1990. x, 212 p. Cloth $27.50.

MURPHY, JOHN W.: *Postmodern Social Analysis and Criticism.* New York: Greenwood, 1989. viii, 177 p. Cloth $39.95.

NERLICH, GRAHAM: *Values and Valuing: Speculations on the Ethical Life of Persons.* New York: Oxford, 1989. 217 p. Cloth $45.00.

NOONAN, HAROLD W.: *Personal Identity.* New York: Routledge, 1989. ix, 262 p. Cloth $55.00.

O'NEILL, JOHN: *The Communicative Body: Studies in Communicative Philosophy, Politics, and Sociology.* Evanston: Northwestern, 1989. xii, 264 p. Cloth $42.95, paper $12.95.

ORMISTON, GAYLE L., and RAPHAEL SASSOWER: *Narrative Experiments: The Discursive Authority of Science and Technology.* Minneapolis: Minnesota UP, 1989. xi, 155 p. Cloth $35.00, paper $14.95.

PAREKH, BHIKHU: *Ghandi's Political Philosophy: A Critical Examination.* Notre Dame: University Press, 1989. ix, 248 p. Cloth $26.95.

PARISH, RICHARD: *Pascal's Lettres Provincialies: A Study in Polemic.* New York: Oxford, 1989. xviii, 203 p. Cloth $49.95.

PARSONS, KEITH M.: *God and the Burden of Proof: Plantiga, Swinburne, and the Analytic Defense of Theism.* Buffalo: Prometheus, 1989. 156 p. Cloth $34.95.

POGGE, THOMAS W.: *Realizing Rawls.* Ithaca: Cornell, 1989. xii, 296 p. Cloth $36.95, paper $12.95.

POLLOCK, JOHN: *How to Build a Person: A Prolegomenon.* Cambridge: MIT, 1990. xi, 189 p. Cloth $22.50.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

PRELLI, LAWRENCE J.: *A Rhetoric of Science: Inventing Scientific Discourse.* Columbia: South Carolina UP, 1989. 320 p. Cloth $29.95.

PRIMORATZ, IGOR: *Justifying Legal Punishment.* Atlantic Highlands, NJ: Humanities, 1989. x, 196 p. Cloth $39.95.

QUINE, W. V.: *Pursuit of Truth.* Cambridge: Harvard, 1990. x, 113 p. Cloth $17.50.

RAPAPORT, HERMAN: *Heidegger & Derrida: Reflections on Time and Language.* Lincoln: Nebraska UP, 1989. 293 p. Cloth $30.00.

REESOR, MARGARET E.: *The Nature of Man in Early Stoic Philosophy.* New York: St. Martin's, 1989. ix, 179 p. Cloth $45.00.

REEVE, C. D. C.: *Socrates in the* Apology. Indianapolis: Hackett, 1989. xv, 207 p. Cloth $24.50.

REIMAN, JEFFERY: *Justice and Modern Moral Philosophy.* New Haven: Yale, 1990. xiv, 322 p. Cloth $30.00.

RESCHER, NICHOLAS: *Cognitive Economy: The Economic Dimension of the Theory of Knowledge.* Pittsburgh: University Press, 1989. x, 168 p. Cloth $29.95.

ROBINSON, DANIEL N.: *Aristotle's Psychology.* New York: Columbia, 1989. xi, 144 p. Cloth $28.00.

ROMANYSHYN, ROBERT D.: *Technology as Symptom and Dream.* New York: Routledge, 1989. xiv, 254 p. Cloth $45.00, paper $16.95.

ROSS, STEPHEN DAVID: *Inexhaustibility and Human Being: An Essay on Locality.* New York: Fordham, 1989. xiii, 330 p. Cloth $40.00.

ROY, SUBROTO: *Philosophy of Economics: On the Scope of Reason in Economic Inquiry.* New York: Routledge, 1989. ix, 236 p. Cloth $37.50.

SACKS, MARK: *The World We Found: The Limits of Ontological Talk.* La Salle: Open Court, 1989. x, 198 p. Cloth $36.95, paper $17.95.

SALMON, WESLEY: *Four Decades of Scientific Explanation.* Minneapolis: Minnesota UP, 1989. xiv, 234 p. Paper $14.95.

SANFORD, DAVID H.: *If P Then Q: Conditionals and the Foundations of Reasoning.* New York: Routledge, 1989. ix, 265 p. Cloth $35.00.

SAUGSTAD, PER: *Language: A Theory of Its Structure and Use.* Atlantic Highlands, NJ: Humanities, 1989. 156 p. Cloth $45.00.

SCHUMACHER, JOHN A.: *Human Posture: The Nature of Inquiry.* Albany: SUNY, 1989. viii, 259 p. Cloth $39.50, paper $12.95.

SHERMAN, NANCY: *The Fabric of Character: Aristotle's Theory of Virtue.* New York: Oxford, 1989. xiv, 213 p. Cloth $39.95.

SHEROVER, CHARLES M.: *Time, Freedom, and the Common Good.* Albany: SUNY, 1989. xiii, 314 p. Cloth $59.50, paper $19.95.

SKORUPSKI, JOHN: *John Stuart Mill.* New York: Routledge, 1989. xv, 431 p. Cloth $59.95.

SMITH, DAVID WOODRUFF: *The Circle of Acquaintance: Perception, Consciousness, and Empathy.* Boston: Kluwer, 1989. xvi, 252 p. Cloth $59.00.

STUMP, ELEONORE: *Dialectic and Its Place in the Development of Medieval Logic.* Ithaca: Cornell, 1989. x, 274 p. Cloth $34.50.

SULLIVAN, ROBERT R.: *Political Hermeneutics: The Early Thinking of Hans-*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Georg Gadamer. University Park: Pennsylvania State UP, 1989. Cloth $22.50.

SUTER, RONALD: *Interpreting Wittgenstein: A Cloud of Philosophy, A Drop of Grammar.* Philadelphia: Temple, 1989. xvi, 256 p. Cloth $34.95.

TAYLOR, MARK C.: *Tears.* Albany: SUNY, 1989. 263 p. Cloth $54.50, paper $17.95.

THOMAS, LAURENCE: *Living Morally: A Psychology of Moral Character.* Philadelphia: Temple, 1989. xii, 260 p. Cloth $29.95.

TILES, MARY: *The Philosophy of Set Theory: An Introduction to Cantor's Paradise.* New York: Blackwell, 1989. viii, 239 p. Cloth $49.95.

VAN FRAASSEN, BAS C.: *Laws and Symmetry.* New York: Oxford, 1989. xv, 395 p. Cloth $55.00, paper $14.95.

VELKEY, RICHARD L.: *Freedom and the End of Reason: On the Moral Foundation of Kant's Critical Philosophy.* Chicago: University Press, 1989. xxi, 222 p. Cloth $29.95.

VELLEMAN, J. DAVID: *Practical Reflection.* Princeton: University Press, 1989. x, 332 p. Cloth $32.50.

WALTON, DOUGLAS N.: *Question-Reply Argumentation.* Westport, CT: Greenwood, 1989. xvi, 408 p. Cloth $49.95.

WARNER, MARTIN: *Philosophical Finesse: Studies in the Art of Rational Persuasion.* New York: Oxford, 1989. x, 401 p. Cloth $64.00.

WARTENBERG, THOMAS E.: *The Forms of Power: From Domination to Transformation.* Philadelphia: Temple, 1990. x, 253 p. Cloth $34.95.

WHITE, DAVID A.: *Myth and Metaphysics in Plato's* Phaedo. Toronto: Associated UP, 1990. 316 p. Cloth $45.00.

WHITLOCK BLUNDELL, MARY: *Helping Friends and Harming Enemies: A Study in Sophocles and Greek Ethics.* New York: Cambridge, 1989. xii, 298 p. Cloth $47.50.

WIERENGA, EDWARD R.: *The Nature of God: An Inquiry into Divine Attributes.* Ithaca: Cornell, 1989. xii, 238 p. Cloth $28.00.

WILLIAMS, HOWARD: *Hegel, Heraclitus and Marx's Dialectic.* New York: St. Martin's, 1989. xv, 256 p. Cloth $35.00.

WILSON, CATHERINE: *Leibniz's Metaphysics: A Historical and Comparative Study.* Princeton: University Press, 1989. x, 350 p. Cloth $39.50.

WINCH, PETER: *Simone Weil: "The Just Balance."* New York: Cambridge, 1989. viii, 234 p. Cloth $44.50, paper $14.95.

WINKLER, KENNETH P.: *Berkeley: An Interpretation.* New York: Oxford, 1989. xiv, 317 p. Cloth $59.00.

XENOS, NICHOLAS: *Scarcity and Modernity.* New York: Routledge, 1989. x, 122 p. Cloth $39.95, paper $12.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVII, NO. 5, MAY 1990

SECOND THOUGHTS ABOUT CHURCH'S THESIS AND MATHEMATICAL PROOFS

I intend to renounce the standard views concerning the nature of Church's thesis. In order to put things in proper perspective, I shall give a brief introduction to the relevant notions and a resumé of the history of the thesis.

The central notion is that of *algorithm*. An algorithm is an effective and completely specified procedure for solving a whole class of problems. Examples are the well-known algorithms for adding and multiplying any two natural numbers. In logic, the truth-table technique of testing a statement form to see whether it is a tautology is an algorithm. An algorithm does not require ingenuity; its application is prescribed in advance and does not depend upon any empirical or random factors.

An algorithm that is applied to natural numbers and yields a natural number as a value is called an *effectively computable function*. In general, the objects that an algorithm is applied to and the objects that it yields can be represented by natural numbers. Therefore, for the sake of simplicity, it is customary to confine attention to effectively computable functions.

It is crucial to understand certain things about effectively computable functions. First of all, we do not mean actual human computability or empirically feasible computability. A function can be computable even though the instructions for it may require more symbols than there are atoms in the universe. Likewise, a value of an effectively computable function may require more time for its computation than the entire past and future history of the human race. When we talk about computability, we ignore any limitations of space, time, or resources. Second, functions are treated extensionally. A function is determined by the ordered pairs of its arguments

0022-362X/90/8705/225–233

© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

and values, not by a particular way in which it is defined. For example, consider the following function:

$$f(x) = \begin{cases} 1 & \text{if Fermat's Last Theorem is true} \\ 0 & \text{if Fermat's Last Theorem is false} \end{cases}$$

This function is an effectively computable function. It is either the constant function 1 (which is effectively computable) or it is the constant function 0 (which is also effectively computable). At the present time, we happen not to know how to compute any values of $f(x)$. Thus, we are adopting a completely classical, nonintuitionistic stance.[1]

Many concrete algorithms have long been known in the history of mathematics and logic. In all such cases, the fact that the alleged algorithm actually is an algorithm is an intuitively obvious fact. (For example, it was so easy to see that our method of adding two natural numbers is an algorithm that we never bothered to prove it. Therefore, it never entered our mind that the notion of algorithm required any further elucidation.) This was not felt to be a difficulty until mathematicians and logicians began to encounter important classes of problems for which there seemed to be no algorithmic solutions. (Examples: (a) Is a given first-order formula logically valid? (b) Is a given arithmetic statement derivable from the axioms of Peano arithmetic? (c) Is a given mathematical statement true? (d) Is a given equation true in a given finitely presented group?) To show the nonexistence of algorithms for solving these problems seemed to call for a mathematically precise equivalent of the notions of algorithm or effectively computable function. The "pressure" of these problems finally became so great that in the middle of the 1930s several quite different explications were independently proposed for the notion of effectively computable function. These proposals were all published in the same year, 1936, by Alonzo Church, Emil Post, and Alan Turing.[2] It turned out that they (and many other later ones) are mutually equivalent. Let us look at Turing's approach, which is the most convincing.

[1] It was from this classical viewpoint that Church's thesis was originally proposed and discussed, and it is still commonly used with such an interpretation. By keeping to this classical setting, I am not denying the tenability of a constructivist philosophy of mathematics.

[2] Church, "An Unsolvable Problem of Elementary Number Theory," *Amer. J. of Mathematics,* LVIII (1936): 345–363; Post, "Finite Combinatory Processes. Formulation I," *Journal of Symbolic Logic,* I (1936): 103–5; Turing, "On Computable Numbers with Applications to the *Entscheidungsproblem*," *Proceedings of the London Mathematical Society,* XLII (1936): 230–265.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Turing wanted to capture the essence of computability by setting up mathematical "models" that would generate all functions, and only such functions, that are computable. These models are what we now call *Turing machines*. When calculating we require a field of computation, such as a piece of paper or a blackboard. A Turing machine uses for this purpose a tape, divided into squares. At any one moment, this tape consists of finitely many squares, but, when necessary, additional squares can be added to the right and left. At any given moment, the squares of the tape may be occupied by certain symbols, at most one symbol to a square. These symbols are chosen from a finite set of symbols, the *alphabet* of the machine. The Turing machine has a *reading device* or *scanner* which, at any given moment, is observing the content of one of the squares of the tape. At any one moment, the scanner can be in any one of a fixed finite collection of *internal states*. (These states correspond to the presumably finite number of states that characterize the structure of a person or computing machine at any given time.) The Turing machine also has a *program*, a finite list of instructions that tell it what to do under all circumstances. If the scanner is in a certain state and is observing a certain symbol, these instructions tell it to do one of four things: (1) replace the observed symbol by another (possibly the same) specific symbol; (2) move one square to the right; (3) move one square to the left; (4) stop. After the appropriate action is performed, the new internal state of the scanner is determined by the program. A Turing machine computes a function $f(x_1, \ldots, x_n)$ of n arguments in the following manner. At the beginning of the computation, the n arguments are written on the tape, separated by blank squares. (Each argument, a natural number m, is represented by a sequence of $m + 1$ strokes, $|||\ldots|$, written in consecutive squares.) The machine starts off in a special state, called the *initial state*, and proceeds to operate according to its instructions. If the machine stops and the only thing left on the tape is the representation of a certain number k, then k is taken to be the value $f(x_1, \ldots, x_n)$. A function computable by some Turing machine is called a *Turing-computable function*. Such a function can be *partial*, that is, it need not be defined for all n-tuples.

Notice that Turing made several normalizing assumptions. Although we usually use at least a two-dimensional field of computation, he restricts the computation to a one-dimensional tape. This really is not an essential limitation, since it is well-known how to encode higher dimensional arrays of symbols in terms of a string of symbols. Second, although we usually move freely around our field of computation, Turing restricts his machine to movements to an

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

adjacent square. Here, one has to verify that this is not a serious limitation; the effect of a "jump" from a square to a nonadjacent square can be accomplished by a suitable sequence of movements to adjacent squares. Third, the fact that the alphabet is finite is not a problem. Actual computing systems use finite alphabets; moreover, a denumerable alphabet $a_1, a_2, \ldots$ can be replaced by a two-symbol alphabet $\{a, '\}$ by letting a' stand for a_1, a'' for a_2, and so on. Fourth, the restriction to a finite set of internal states seems to be acceptable. Turing justifies this as follows: "If we admitted an infinity of states of mind, some of them will be 'arbitrarily close' and will be confused." (If there is any weak link in Turing's presentation, this is probably it.) Church's thesis can now be formulated.

> Church's thesis (CT): A function is effectively computable if and only if it is Turing-computable.

The Turing-computable functions have been shown to be identical with the so-called *partial recursive functions* (defined by S. C. Kleene[3]). Church's thesis is often stated in terms of partial recursive functions instead of Turing-computable functions, because the former are easier to work with from a technical standpoint.

Here are some of the arguments in favor of CT.

(a) The host of notions proposed as equivalents of effectively computable function have all turned out to be equivalent. This seems to show that the underlying intuitive target notion has been hit.[4]
(b) It is easy to see that one half of CT is true: every Turing-computable function is effectively computable.
(c) The other half of CT, that every effectively computable function is Turing-computable, has been confirmed for the very large number of cases in which it has been tested.
(d) The arguments given by Turing show that the special methods of computation used by Turing machines involve no essential restriction on the general methods of computation employed by (idealized) human beings or computers.

CT is now almost universally accepted by mathematicians and logicians. In the theory of recursive functions, it is constantly resorted to without explicit mention.[5] Moreover, all the limitative theorems of modern logic, such as Gödel's incompleteness theorem, depend for their significance on the validity of CT. For example, Church's theorem states that there is no recursive decision procedure for

[3] *Introduction to Metamathematics* (Princeton: Van Nostrand, 1952).
[4] Of course, this is not conclusive. It is conceivable that all the equivalent notions define a concept that is related to, but not identical with, effective computability.
[5] See the standard reference work by Hartley Rogers, Jr., *Theory of Recursive Functions and Effective Computability* (New York: McGraw-Hill, 1967).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

first-order logical validity.[6] It is CT that then allows us to draw the important conclusion that there is no effective procedure for deciding whether a first-order formula is logically valid.

CT is generally regarded as an assertion of the extensional equivalence of the notions of effectively computable function and partial recursive function (or Turing-computable function). Because the former notion is intuitive and the latter is alleged to be a precise mathematical concept, it is further asserted that it is impossible to *prove* CT; hence, CT is just a conjecture or a thesis, and is not susceptible of a rigorous proof. This is what almost everybody believes about CT; for example, Church and Kleene:

> This definition is thought to be justified by the considerations which follow, so far as positive justification can ever be obtained for the selection of a formal definition to correspond to an intuitive notion (Church, *op. cit.*, §7).

> Since our original notion of effective calculability of a function is a somewhat vague intuitive one, [CT] cannot be proved . . . While we cannot prove [CT], since its role is to delimit precisely a hitherto vaguely conceived totality, we require evidence that it cannot conflict with the intuitive notion which it is supposed to complete; i.e. we require evidence that every particular function which our intuitive notion would authenticate as effectively calculable is [recursive]. The thesis may be considered a hypothesis about the intuitive notion of effective calculability, or a mathematical definition of effective calculability; in the latter case, the evidence is required to give the theory based on the definition the intended significance (Kleene, *op. cit.*, pp. 317–9).

These standard views about CT hold it to be a *rational reconstruction*, in the sense of Rudolf Carnap and Carl Hempel. A rational reconstruction is a precise, scientific concept that is offered as an equivalent of a prescientific, intuitive, imprecise notion. In all standard cases in which the original intuitive notion is definitely known to apply or not to apply, the rational reconstruction should yield the same outcome. The rational reconstruction may go beyond the original notion, however, in cases where the latter notion is not determinate. Moreover, the rational reconstruction need not, and usually does not, have the same psychological effect as the original notion; expert users of the language generally do not think of the rational reconstruction when they use the original notion. Confirmation of the correctness of a rational reconstruction apparently must involve, at least in part, an empirical investigation. The correctness of the reconstruction cannot be proved.

[6] More precisely, if f is the function such that $f(n) = 1$ if n is the Gödel number of a logically valid first-order formula and $f(n) = 0$ otherwise, then f is not recursive.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

I would like to challenge the standard conception of CT as an unprovable thesis. My viewpoint can be brought out clearly by arguing that CT is another in a long line of well-accepted mathematical and logical "theses," and that CT may be just as deserving of acceptance as those other theses. Of course, these theses are not ordinarily called "theses," and that is just my point.[7]

The following are a few obvious cases of "theses" that have never been called "theses," but probably merit this term as much or as little as CT.

I. *Functions.* In contemporary logic and mathematics, a function f is defined to be a set of ordered pairs satisfying the following condition: (*) If $(x, y) \in f$ and $(x, z) \in f$, then $y = z$. In classical mathematics, a function was tied to a rule for calculating it, generally by means of a formula of a standard type. In the course of the nineteenth century, mathematicians reluctantly came to the conclusion that this conception of function was too narrow; now, with the exception of constructivists, everyone accepts the definition given above. The earliest source that I can find that contains this definition is: G. Peano, "Sulla definizione di funzione," *Atti della Reale Accademia dei Lincei*, Classe di scienze fisiche, matematiche e naturali, XX (1911): 3–5. Hence, we might call the identification of this definition with the intuitive notion of function "Peano's thesis." It is clear that an intuitively understood function determines a unique set of ordered pairs satisfying condition (*). Conversely, such a set f can be thought of as determining a rule for calculating a function: to each x assign that object y such that $(x, y) \in f$. (Since this "rule" is not necessarily constructively definable, constructivists would reject this part of Peano's thesis.) It appears that, in contradistinction to CT, Peano's thesis has no important mathematical consequences and is useful only for foundational purposes: (a) in an axiomatic presentation of mathematics, we do not need to assume the concept of function as primitive; (b) together with other axioms of ZF (Zermelo-Fraenkel set theory), this definition allows us to prove that, for any sets A and B, the set A^B of all functions from B into A exists.

II. *Tarski's definition of truth.* Consider a first-order language L

[7] A search of the literature has turned up only one paper with explicit references to 'theses' in the history of logic and mathematics: Stewart Shapiro, "Understanding Church's Thesis," *Journal of Philosophical Logic*, x (1981): 353–365. Shapiro seems to accept the standard view of the inherent unprovability of Church's thesis, however. While preparing the final form of the present paper, another such reference came to my attention: R. L. Epstein & W. A. Carnielli, *Computability: Computable Functions, Logic, and the Foundations of Mathematics* (Pacific Grove, CA: Wadsworth, 1989), ch. 25. They consider theses concerning the notions of "circle" and "continuous function." But they also accept the orthodox view of CT: "(CT) is a nonmathematical thesis: it equates an intuitive notion . . . with a precise, formal one" (p. 86).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

and a structure $\mathcal{M}$ based on this language. For any sentence $\mathcal{B}$ of L, we understand what it means for $\mathcal{B}$ to be true in $\mathcal{M}$. In the 1930s, Tarski gave a set-theoretic definition of '$\mathcal{B}$ is true in $\mathcal{M}$' that is now universally acknowledged to be equivalent to the corresponding intuitive notion. By mathematical induction, using some basic properties of the intuitive notion and of Tarski's notion, we can convince ourselves of the validity of "Tarski's thesis." Like the notion of recursiveness, Tarski's definition has found important applications, particularly in connection with proofs of consistency and independence.

III. *Logical validity.* The currently accepted semantical or model-theoretic definition of logical validity is that a first-order sentence is logically valid if it is true in all structures. (Notice that this definition attains greater precision if we understand 'true' in the Tarskian sense.) This definition probably goes back to the 1920s, although I cannot identify its exact origin. In any case, Gottlob Frege, Bertrand Russell, A. N. Whitehead, and other earlier logicians certainly understood the notion of logical validity, although it is highly unlikely that they had thought of, or would have assented to, the model-theoretic definition. (Frege thought of his individual variables as ranging over arbitrary objects in an unrestricted universe, and predicate letters as ranging over arbitrary properties and relations on the universe. Hence, his conception of logical validity, if he had ever formulated it, probably would have differed from the model-theoretic definition.) Thus, we are dealing here again with a thesis, one that asserts the equivalence of the model-theoretic definition with a previous intuitive notion. There are plausible arguments in favor of the thesis. A theorem of model theory tells us that a sentence that is true for all structures of a given cardinality α is also true for all structures of cardinality less than α. Hence, if we think (at our peril) of α as the cardinality of the universe, then Frege's conception of logical validity would imply the model-theoretic version. Conversely, if a sentence $\mathcal{B}$ is true for all structures, then Gödel's "completeness theorem" implies that $\mathcal{B}$ is provable in the first-order predicate calculus. But all the axioms of that calculus would be universally accepted as logically valid and its rules of inference would be universally acknowledged to lead from logically valid formulas to logically valid formulas. Hence, $\mathcal{B}$ should be accepted as logically valid in the Fregean (or any other) sense.

IV. *Limits.* The intuitive notion of limit was widely used in mathematical analysis in the eighteenth century, and, in the first half of the nineteenth century, A. L. Cauchy showed how limits could be used to give a firm foundation to the basic ideas of calculus, such as continuity and differentiation. It seems to have been only in the 1860s,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

however, that Karl Weierstrass was the first one to give the well-known ϵ-δ type of definition of a limit of a function (and a corresponding definition for a limit of a sequence). The Weierstrassian definitions can be seen to capture exactly the meaning of the preceding intuitive notions; the "Weierstrass thesis" ranks among the most firmly established mathematical theses. When we ordinarily work with limits and when we introduce the idea to students, we still rely on the original intuitive picture, but we resort to the more precise definition in difficult cases.

These four examples form just a small sample of a large body of logical and mathematical theses. A few other examples are the notion of measure as an explication of area and volume, the definition of dimension in topology, the definition of velocity as a derivative, the definition of logical implication and logical equivalence in first-order logic, and the definitions of circle, triangle, interior of an angle, and many other geometric concepts.

Now that we have other theses with which to compare CT, here is the main conclusion I wish to draw: it is completely unwarranted to say that CT is unprovable just because it states an equivalence between a vague, imprecise notion (effectively computable function) and a precise mathematical notion (partial-recursive function). My argument is based on three points.

(1) The concepts and assumptions that support the notion of partial-recursive function are, in an essential way, no less vague and imprecise than the notion of effectively computable function; the former are just more familiar and are part of a respectable theory with connections to other parts of logic and mathematics. (The notion of effectively computable function could have been incorporated into an axiomatic presentation of classical mathematics, but the acceptance of CT made this unnecessary.) The same point applies to theses I–III. Functions are defined in terms of sets, but the concept of set is no clearer than that of function and a foundation of mathematics can be based on a theory using function as primitive notion instead of set. Tarski's definition of truth is formulated in set-theoretic terms, but the notion of set is no clearer than that of truth. The model-theoretic definition of logical validity is based ultimately on set theory, the foundations of which are no clearer than our intuitive understanding of logical validity.

(2) The assumption that a proof connecting intuitive and precise mathematical notions is impossible is patently false. In fact, half of CT (the "easier" half), the assertion that all partial-recursive functions are effectively computable, is acknowledged to be obvious in all textbooks in recursion theory. A straightforward argument can be

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

given for it. (The so-called initial functions are clearly effectively computable; we can describe simple procedures to compute them. Moreover, the operations of substitution and recursion and the least-number operator lead from effectively computable functions to effectively computable functions. In each case, we can describe procedures that will compute the new functions.) This simple argument is as clear a proof as I have seen in mathematics, and it is a proof in spite of the fact that it involves the intuitive notion of effective computability. The fact that it is not a proof in ZF or some other axiomatic system is no drawback; it just shows that there is more to mathematics than appears in ZF.

(3) Another difficulty with the usual viewpoint concerning CT is that it assumes that the only way to ascertain the truth of the equivalence asserted in CT is to *prove* it. In mathematics and logic, proof is not the only way in which a statement comes to be accepted as true. Of course, this is a consequence of the truism that not all truths can be proved; proofs must assume certain axioms and rules of inference. But, over and above this, we have observed in our four examples above that equivalences between intuitive notions and apparently more precise mathematical notions often are simply "seen" to be true without proof, or are based on arguments that are a mixture of such intuitive perceptions and standard logical and mathematical reasoning. (Notice also that these perceptions and arguments seem to be of a nonempirical nature.)

That CT is true follows, I believe, from Turing's analysis of the essential elements involved in computation. (Further, even more sophisticated, Turing-like analyses have been given.[8]) But this is not what I have tried to establish. The point I have attempted to make is that equivalences between intuitive notions and "precise" notions need not always be considered unprovable theses. Such equivalences sometimes can be directly perceived to be valid, sometimes they can be justified by a combination of directly perceived truths and logical argument, and, in other cases, it is possible to discover entirely convincing proofs.[9]

ELLIOTT MENDELSON
City University of New York/Queens College

[8] A.N. Kolmogorov and V.A. Uspenski, "On the definition of an algorithm," *Uspekhi Mat. Nauk*, VIII (1953): 125–176 (*Amer. Math. Soc. Translations*, XXIX (1963): 217–245). R. Gandy, "Church's Thesis and Principles for Mechanisms," *The Kleene Symposium* (Amsterdam: North-Holland, 1980), pp. 123–148.

[9] After completion of this paper, it came to my attention that G. Kreisel had made a somewhat similar point in his paper, "Informal Rigour and Completeness Proofs," *Problems in the Philosophy of Mathematics*, I. Lakatos, ed. (Amsterdam: North-Holland, 1967), pp. 138–186, esp. p. 176.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

SECOND-ORDER LOGIC, FOUNDATIONS, AND RULES*

> Not empiricism and yet realism in philosophy, that is the hardest thing.
>
> And the picture that might occur to someone here is that of a short bit of handrail, by means of which I am to let myself be guided further than the rail reaches. [But there *is* nothing there; but there isn't *nothing* there!]
>
> The difficult thing here is not, to dig down to the ground; no, it is to recognize the ground that lies before us as the ground.
>
> Ludwig Wittgenstein[1]

AN interpreted, or partially interpreted, language is called "second-order" or "higher-order" if it has variables that range over relations, propositional functions, properties, classes, or sets of whatever is in the range of the ordinary, or first-order variables. Many studies have paid careful attention to the differences among these items. It is said, for example, that propositional functions are intensional, while sets and classes are extensional; sets are "constituted by their elements," while classes are extensions of properties.[2] Here I gloss over these differences and use the terms interchangeably. If an identity relation on such items is needed, I take them to be extensional, but few, if any, of my remarks depend on this. I apologize to sensitive readers.

In any case, the distinguishing feature of second- or higher-order languages is not so much the nature of the individual items that fall in the range of the extra variables, but rather their extension or *totality*. It is often noted along these lines that the central feature of second-order logic is its semantics, as opposed to the language itself and its deductive system. Indeed, one may go so far as to assert that

* I would like to thank the participants of the Ohio University Conference, especially John Corcoran, Richard Butrick, and George Boolos, for many hours of fruitful discussion. I am also indebted to the members of a discussion group on the Skolem paradox—Beth Cohen, Rick DeWitt, Robert Kraut, Ronald Laymon, Barbara Scholz, and George Schumm—and to my colleagues David McCarty and Allan Silverman.

[1] *Remarks on the Foundations of Mathematics* (Cambridge: MIT, 1978, Rev. ed.), pp. 325, 430, 333.

[2] The topic is treated and illuminated throughout Charles Parsons, *Mathematics in Philosophy* (Ithaca: Cornell, 1983); see esp. Essay 8, "Sets and Classes."

0022-362X/90/8705/234–261 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

languages and deductive systems, by themselves, are neither first-order nor higher-order. I have said this myself, for example.[3] I still think it is correct, especially when one focuses on formal or semi-formal Tarskian model theory, but it can be, for this reason, misleading. It depends on what one takes "semantics" to be.

I show here that the major issues surrounding second-order languages re-emerge when one compares theories of formal proof with informal or preformal mathematical practice. In a sense, this is "semantics" also, since issues of "meaning" are involved, but it is not traditional model theory.

I

Let L2 be a second-order language, in which, for simplicity, all second-order variables are monadic. Let L1 be the first-order restriction of L2. The range of the first-order variables, or the *domain,* may be unspecified, as in formal logic, or there may be a structure, such as the natural numbers, the real numbers, or the set-theoretic hierarchy, that serves as *intended domain.*

As usual, first-order variables are lower-case letters and predicate variables are upper case letters. A deductive system for L2 can be obtained from a deductive system for L1 by adding straightforward extensions of the quantifier axioms (e.g., $\forall X(\Phi(X)) \rightarrow \Phi(Y)$), and the axiom scheme of comprehension:

$$\exists X \forall x(Xx \equiv \Phi(x))$$

for each formula Φ not containing X free.[4]

A brief sketch of the model theory will help focus attention on the central issues. There are (at least) three semantics that have been introduced for languages like L2, but two of these are "equivalent" in a straightforward sense.

In *standard semantics,* which makes the logic properly second-order, the predicate variables range over the collection of all subsets of the domain. With a little more detail, a *standard interpretation* of

[3] See my "Second-order Languages and Mathematical Practice," *The Journal of Symbolic Logic,* L (1985): 714–742; also my *Foundations without Foundationalism: A Case for Second-order Logic* (New York: Oxford, forthcoming).

[4] See David Hilbert and Wilhelm Ackermann, *Grundzüge der theoritischen Logik* (Berlin: Springer, 1928). Notice that the formula Φ in the comprehension scheme may contain bound-predicate variables. Thus, some instances are impredicative. To avoid this, restricted versions of the scheme have been considered, resulting in ramified type theory. On the other hand, if function and relation variables are introduced into the language, the axiom of choice may be included:

$$\forall R(\forall x \exists y Rxy \rightarrow \exists f \forall x Rxfx)$$

A symbol for identity-between-predicates can also be introduced, together with the axiom of extensionality:

$$\forall X \forall Y[\forall x(Xx \equiv Yx) \rightarrow X = Y]$$

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

L2 is a structure $\langle d, I \rangle$, in which d is the domain and I an appropriate assignment on d to the common nonlogical terminology. A *variable assignment* consists of a function from the collection of first-order variables to d and a function from the collection of predicate variables to the powerset of d. The relation of *satisfaction,* between interpretations, variable assignments, and formulas is defined in the straightforward way.

Although this semantics is sound for all common deductive systems, the completeness, compactness, and Löwenheim-Skolem theorems all fail; and categorical characterizations of infinite structures are possible.[5]

Notice that a standard interpretation for L2 is exactly the same as an interpretation for the first-order L1, namely, a domain and an assignment to the (common) nonlogical terminology. That is, in standard semantics, by fixing a domain, one thereby fixes the range of both the first-order and the second-order variables. There is no further "interpreting" to be done. This is not the case with the next two candidate semantics. In both cases, one must separately determine a range for the first-order variables and a range for the second-order variables. That is the crucial difference.

In *Henkin semantics,* the predicate variables range over a fixed collection of subsets of the domain (which may not include all of the subsets). That is, a *Henkin interpretation* of L2 is a structure $\langle d, D, I \rangle$, in which d is a domain, D a subset of the powerset of d, and I an appropriate assignment on d to the nonlogical terminology. Informally, D is the range of the predicate variables. A *Henkin variable assignment* consists of a function from the collection of first-order variables to d and a function from the collection of predicate variables to D. Again, the relevant satisfaction relation is straightforward.

Ironically, this semantics is not sound for many deductive systems. The reason is that some Henkin interpretations do not satisfy the comprehension scheme (and other axioms that may have been added). Define an interpretation to be *faithful* to a given deductive system for L2 if it satisfies its second-order axioms. Leon Henkin[6] established a completeness theorem for L2 with this semantics: a sentence Φ is provable in a given deductive system for L2 if and only if Φ is satisfied by every Henkin interpretation faithful to that deductive system. The compactness and Löwenheim-Skolem theorems

[5] See George Boolos and Richard C. Jeffrey, *Computability and Logic* (New York: Cambridge, 1980, 2nd ed.), ch. 18; and my *Foundations without Foundationalism.*

[6] "Completeness in the Theory of Types," *Journal of Symbolic Logic,* XV (1950): 81–91.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

are also provable for L2 with (faithful) Henkin semantics. Consequently, there are no categorical characterizations of infinite structures.

It is straightforward to see that a standard interpretation is equivalent to the Henkin interpretation in which the designated range of the predicate variables is the entire powerset of the domain. These "full interpretations" are faithful to all common deductive systems.

In first-order semantics, L2 is regarded as a *two-sorted* first-order language, and the "predication" or "membership" relation between properties (or classes) and objects, is regarded as nonlogical. In particular, a first-order interpretation of L2 is a structure $\langle d_1, d_2, \langle I, p\rangle\rangle$, in which d_1 and d_2 are sets, p is a subset of $d_1 x d_2$, and I is an appropriate assignment (on d_1) to the (other) nonlogical terminology. Informally, d_1 is the range of the first-order variables, d_2 is the range of the predicate variables, and p is the interpretation of the "predication" relation. A *first-order variable assignment* s consists of a function from the collection of first-order variables to d_1 and a function from the collection of predicate variables to d_2. Once again, the notion of satisfaction is straightforward. One clause is:

> Let s assign u (an element of d_1) to the variable x and v (an element of d_2) to X. Then $\langle d_1, d_2, \langle p, I\rangle\rangle$ and s satisfy the atomic formula Xx if and only if the pair $\langle u, v\rangle$ is in p.

As above, define a structure to be *faithful* to a deductive system if it satisfies each second-order axiom thereof. Of course, the completeness, compactness, and Löwenheim-Skolem theorems all hold for this semantics (restricted to faithful interpretations).

Every Henkin interpretation $\langle d, D, I\rangle$ is equivalent to the first-order interpretation in which the "second" domain is D and p is the "real" membership or predication relation. A converse (of sorts) holds as well: for every first-order interpretation M^1, there is a Henkin interpretation M^H, whose domain is the first domain of M^1, such that for each sentence Φ of L2, $M^H \models \Phi$ iff $M^1 \models \Phi$.[7]

Let us return to the difference between first-order logic and second-order logic. Clearly, there is nothing to prevent a first-order language from being "about" predicates, properties, sets, or any-

[7] See Paul Gilmore, "The Monadic Theory of Types in the Lower-predicate Calculus" (*Summaries of Talks Presented at the Summer Institute of Symbolic Logic at Cornell*, Institute for Defense Analysis, 1957). For each $A \in d_2$ let $S(A)$ be the set $\{a \in d_1 \mid \langle a, A\rangle \in p\}$. Let $D^1 = \{S(A) \mid A \in d_2\}$. Then the Henkin interpretation equivalent to M^1 is $\langle d_1, D^1, I\rangle$. To take a frivolous example, suppose that in M^1, d_1 is a collection of apples and d_2 is a collection of oranges. Then in the Henkin model equivalent to M^1, each orange is "replaced" by the set of apples that bear the (interpretation of the) "predication" relation to it.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

thing else for that matter, provided only that such items exist and that one can make coherent statements about them. Similarly, there is nothing to prevent the introduction of a two-sorted first-order language, with one sort of variable intended to range over "objects" and another sort intended to range over "properties" of these objects. The relation of "predication" between properties and objects would then be nonlogical, but I would suggest that determining the distinction between logical and nonlogical terminology is not separate from determining the intended "semantics" of the language.

In a recent conference proceedings, John Corcoran[8] argues that discourse about properties (and relations) is a natural (and inevitable) extension of discourse about "objects," as indicated by the presence of common nouns and straightforward grammatical constructions. In present terminology, the thesis is that languages with variables ranging over properties (and relations) are natural extensions of languages with variables ranging over objects. This is particularly clear, I believe, in the development of formal languages to codify the underlying logics of mathematical systems.

There is a striking affinity between Corcoran's remarks and W. O. Quine's analysis of ontological commitment in language. Both focus on the development of linguistic techniques for acknowledging and referring to objects. Quine, of course, is a steadfast opponent of second-order logic *as logic,* but he is not an opponent of set theory, as mathematics. This indicates that the analysis of language that shows that talk of sets is natural does not thereby determine the totality of the range of the property variables. In the present framework, the Quine-Corcoran analysis does not, by itself, decide among the semantics for the extended languages. Many of Corcoran's arguments in favor of second-order *languages* equally support standard semantics and Henkin, or first-order, semantics.[9]

The proponents of second-order logic *as logic,* then, hold that

[8] "Second-order Logic," in *Proceedings Inference OUIC 86,* D. Moates and R. Butrick, eds. (Athens: Ohio UP, 1987), pp. 7–31.

[9] See Quine, *Word and Object* (Cambridge: MIT, 1960), and, for the attack on second-order logic, Quine, *Philosophy of Logic* (Englewood Cliffs, NJ: Prentice-Hall, 1970). Corcoran's presentation of second-order languages as (natural) *extensions* of first-order ones does not reflect the history of logic (nor was it intended to). First-order languages were first studied [by Leopold Löwenheim, "Über Möglichkeiten im Relativkalkül," *Mathematische Annalen,* LXXVI (1915): 447–479; and Hilbert and Ackermann, *op. cit.*] as distinct *subsystems* of previously developed higher-order logics [see also G. Moore, "Beyond First-order Logic: The Historical Interplay between Logic and Set Theory," *History and Philosophy of Logic,* I (1980): 95–137; and "The Emergence of First-order Logic," in W. Aspray and P. Kitcher, eds., *Essays in the History and Philosophy of Mathematics* (Minneapolis: Minnesota Studies in the Philosophy of Science, 1988), pp. 95–135].

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

second-order terminology with standard semantics is sufficiently clear, intuitive, or unproblematic to serve the underlying framework for axiomatization and foundations of mathematics. The claim is that, once a domain (for the first-order variables) is fixed, there is a clear and unambiguous understanding of such locutions as 'all properties' or 'all subsets' thereof.[10] Such locutions play central roles in foundations of mathematics.

The first-order "opposition," on the other hand, rejects this. Most do not claim that sets do not exist, nor that discourse about sets or properties is somehow illegitimate. Indeed, many of the major proponents of first-order logic, such as Skolem, Quine, and Gödel, have explicitly acknowledged set theory and contributed to its development. Moreover, Quine and Gödel adopted and defended attitudes of realism toward set theory (each in his own way). The claim seems to be that variables ranging over all of the properties of a fixed domain are not sufficiently clear to serve the foundation of mathematics, and that this terminology itself stands in need of further "foundation." That is, the relevant theory of sets ought to be formulated as an axiomatic theory in its own right, presumably a *first-order* theory. Quine calls second-order logic "set theory in disguise."[11] It almost goes without saying that the relevant notion of "sufficient clarity" could use some clarification. The issue here is a difficult one and it seems impossible to avoid begging at least some of the questions.

It might be noted in passing that *foundationalism,* the view that there is, or can be, a unique, secure "foundation" for all knowledge, is almost universally rejected today. Variants that restrict the foundation to mathematical knowledge may have some contemporary advocates, but if these too are rejected, then one should expect

[10] Notice that the foregoing claim is markedly weaker than the assertion that the entire set-theoretic hierarchy is clear and intuitive. Zermelo set theory, for example, has a powerset operator that can be iterated—even into the transfinite. The second-order thesis is only that, once a domain is fixed, the range of the second-order variables is also fixed. Let d be a domain, regarded as unproblematic. The thesis at hand then accepts the notion of *all* subsets of d—one powerset of d. If *that* is now taken as another "domain," the second-order thesis sanctions its powerset—the collection of all sets of subsets of d. But even this "construction" cannot be iterated into the transfinite. At most, the "commitment" is to "n^{th} order" sets (of d) for any natural number n.

[11] Quine, *Philosophy of Logic.* As far as I can determine, neither Skolem nor Gödel ever *argued* in print for the restriction to first-order logic. Gödel's correspondence with Zermelo illustrates this well. See I. Grattan-Guinness, "In Memoriam Kurt Gödel: His 1931 Correspondence with Zermelo on His Incompletability Theorem," *Historia Mathematica,* VI (1979): 294–304; J. Dawson, "Completing the Gödel-Zermelo Correspondence," *Historia Mathematica,* XII (1985): 66–70; and Moore, *op. cit.*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

neither a unique nor an absolutely secure foundation for mathematics. There may be different foundations, perhaps serving different purposes, but, in any case, the goal of absolute certainty is not among them. The present contribution is made in this spirit.[12]

II

Imagine a dialogue between an advocate of second-order logic, call him Second, and an advocate of first-order logic, call her First. Second begins by developing a formal language with variables ranging over properties, and he explains (in informal language) the intuitive, standard "semantics," the idea that the property variables range over all properties, or all sets, of numbers. He then proves that arithmetic and analysis, as formulated in the second-order language, are categorical, and he shows that the real numbers are not countable. At this point, First raises a question concerning the range of the second-order variables. She asserts that the meaning of the second-order terminology is not very clear and she requests that it be developed more fully. Second could retort that First knows perfectly well what locutions like 'all subsets' mean, and he may accuse her of making trouble for the sake of making trouble. They would then be at a standoff.

On the other hand, Second may regard the request for clarification as legitimate. One can, indeed, do formal semantics and, in effect, provide a foundation for the background theory. So Second proceeds to develop a (weak) version of axiomatic set theory sufficient to formulate the standard semantics of the second-order theories. Call the formal metatheory MT. Second shows how the proofs of the categoricity and cardinality theorems can be rendered in MT.

[12] The relationship between foundationalism and foundations of mathematics would make a fascinating study. Skolem's "Some Remarks on the Foundation of Set Theory" [*Proceedings of the International Congress of Mathematicians, Cambridge, Massachusetts* (Providence: American Mathematical Society, 1952): 695–704] presents three goals for foundations of mathematics. One is to "obtain a way of reasoning which is logically correct so that it is clear and certain in advance that . . . what we prove are truths in some sense." This is called a "natural" standpoint, and is identified with logicism. Another goal is to "have a foundation which makes it possible to develop present day mathematics, and which is consistent so far as is known yet." It is lamented that this "practical" or "opportunistic" standpoint has the "unpleasant feature that we can never know when we have finished the foundation of mathematics. We are not only adding new floors at the top of our building, but from time to time it may be necessary to make changes in the basis." The third outlook is the Hilbert program, which is "giving up of the logicistic standpoint and not being content with the opportunistic one." Skolem's preference for foundationalism is clear, but he realized its difficulties. See S. Wagner, "The Rationalist Conception of Logic," *Notre Dame Journal of Formal Logic*, XXVIII (1987): 3–35, for a contemporary discussion of rationalism and foundationalism in logic.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

At this point, things get interesting. First applauds the effort. She agrees that the development of MT is a significant event and clarifies things considerably. She even contributes to the development of MT by proving some theorems and suggesting new axioms. But she takes MT as a first-order theory. In particular, she claims that the semantical theory is itself subject to the Löwenheim-Skolem theorems, and the like. Thus, MT has many interpretations. First argues that the categoricity theorems only show that, within each interpretation of the semantics MT, all models of arithmetic are isomorphic and, similarly, within each interpretation all models of analysis are isomorphic. Also, within each model of the semantics, the interpretation of the real numbers is uncountable *in that model.* Such is relativism.

Second retorts that MT is not to be regarded as just another uninterpreted theory with various models of various cardinalities. On the contrary, MT has an *intended interpretation*—the intuitive semantics of the original languages. In short, the categoricity results should be taken at face value, referring to the possible domains for the original languages, not to this or that model of a formal theory like MT. The results are that arithmetic and analysis are really categorical, not just categorical in each model of MT, and Second insists that the real numbers are really uncountable. Moreover, he points out that the argument that First makes depends in part on the Löwenheim-Skolem theorems, which are demonstrably false for the second-order languages under study. Remember the categoricity theorems.

The reply at this point is predictable. First claims that the "refutations" of the Löwenheim-Skolem theorems only show that, within each interpretation M of MT, each model m of, say, analysis, is uncountable in M. It does not rule out the possibility that this same model m is countable in some more encompassing model of MT (but of course m would not be a model of analysis in the more encompassing structures). Indeed, we cannot rule out the possibility of countable models of MT itself. The theorems only show that the same structure cannot be both uncountable and a model of analysis from the same perspective. As for the assertion that MT is already interpreted, First requests clarification of this "intended semantics for the original languages." Second reminds First that he has already accomplished this. MT is the characterization, and First accepted it. But she seems to insist that MT must itself be interpreted. We seem to have reached another standoff.

Upon further examination (since neither Second nor First are stubborn), it is seen that the dispute centers on the axiom of separation (in MT), which states that, if x is a set and P a property that is

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

well-defined on the members of x, then there is a subset y of x that contains all and only the members of x that have the property P. Its straightforward formulation is:

$$\forall x \exists y \forall z[z \in y \equiv (z \in x \ \& \ Pz)]$$

This, of course, is a second-order sentence. Second insists that it be given a *standard* reading, with the variable P ranging over all properties of the universe (i.e., all classes), while First insists on a first-order (or Henkin) interpretation. At this point, Second can respond to the challenges by developing a semantics for MT, a metasemantics. Call it MMT. But then First would understand MMT as a first-order theory and Second would not. Our desire to avoid (or postpone) the standoff between First and Second has led to a regress.

The skepticism, or relativism, that First advocates is not limited to second-order logic. There is a remarkably similar regress or standoff concerning the range of the first-order variables in rather basic theories. They can be regarded as uninterpreted and subject to a wide variety of standard and nonstandard modeling, or else one can insist that there is an unambiguous, intuitive understanding of the range of the variables. Suppose, for example, that First asks what one means by 'the natural numbers', the intended domain of arithmetic. Second responds with an axiomatization of Peano arithmetic, the crucial item of which is the induction axiom:

$$(P0 \ \& \ \forall x(Px \rightarrow Psx)) \rightarrow \forall x Px$$

First raises the (by now) usual question concerning the range of the property variable P. She takes induction to be a first-order axiom scheme, and points out that the theory, thus construed, has lots of models of various cardinalities. Of course, Second rejects this construal and replies that the understanding he has in mind—the intended model—is the "smallest" of the models. It is the structure that is an initial segment of all of the first-order models. First asks for a characterization of 'smallest model' or 'initial segment of all models', claiming that the quantifiers in these locutions are too problematic to serve foundations without further ado. Perhaps in frustration, Second asserts that by 'natural number', he means "member of the sequence 0, 1, 2, 3, 4, ..." First asks about the '. . .'.

In sum, our advocate of second-order logic holds that reference to, say, the natural numbers is clear and unambiguous, at least now. Any two mathematicians who discuss arithmetic are talking about the same (or isomorphic) structures. Suppose that First and Second agree, at least for the sake of argument, that formal arithmetic adequately codifies pretheoretical discourse—as they agreed that MT

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

codifies informal semantics. Second takes the categoricity results to confirm his prior belief that arithmetic is unambiguous. But the results involve locutions like 'all properties' or 'all subsets', and the "confirmation" depends on these locutions themselves being unambiguous. First insists that there is no reason to hold this for the second-order variables and, thus, no reason to accept the confirmation of the original belief that arithmetic is unequivocal.

I conclude that the skepticism or relativism concerning second-order languages applies (or not) just as well to mathematical theories whose intended domains are infinite. Relativism concerning second-order variables goes with what may be called "object language relativism." On the other hand, one who is willing to accept classical mathematics, more or less as it stands, should also accept second-order languages. No more is involved. Alonzo Church[13] makes a similar point:

> Objections may indeed be made to this new point of view, on the basis of the sort of *absolutism* it presupposes . . . But it should be pointed out that this . . . is already inherent in classical mathematics generally, and it is not made more acute or more doubtful, but only more conspicuous, by its application to [logic]. For our definition of the consequences of a system of postulates . . . can be seen to be not essentially different from [that] required for the . . . treatment of classical mathematics . . . It is true that the non-effective notion of consequence, as we have introduced it . . . presupposes a certain absolute notion of ALL propositional functions of individuals. But this is presupposed also in classical mathematics, especially classical analysis, . . . (*ibid.*, p. 326N).

Parts of these dialogues can be seen in exchanges between historical figures early this century. Ernst Zermelo,[14] who corresponds to Second, formulated axiomatic set theory in 1908. He defined a propositional function $P(x)$ to be *definit* for a domain d if, for each element x of d, "if the fundamental relations of the domain, by means of the axioms and the universally valid laws of logic, determine without arbitrariness whether [P(x)] holds or not." The axiom of separation was stated in these terms:

> Whenever the propositional function [P(x)] is *definit* for all elements of a set M, M possesses a subset containing as elements precisely the elements x of M for which [P(x)] is true.

[13] *Introduction to Mathematical Logic* (Princeton: University Press, 1956).

[14] "Neuer Beweis für die Möglichkeit einer Wohlordnung," *Mathematische Annalen*, LXV (1908): 107–128; and "Untersuchungen über die Grundlagen der Mengenlehre. I," *Mathematische Annalen*, LXV (1908): 261–281. Both of these are translated in J. van Heijenoort, *From Frege to Gödel* (Cambridge: Harvard, 1967).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Much of the extensive criticism that Zermelo's work engendered was directed at the axiom of choice, but separation was not ignored. Among the attempted improvements, the most successful was that of Thoralf Skolem,[15] in 1922. Like our character First, he formulated separation as a scheme, one instance for each formula of the (first-order) language. This, of course, is current practice.

When Zermelo returned to the axiomatization in 1929, he agreed that the original notion of *definit* was not formulated with sufficient precision, and he seemed to acknowledge that the explication depends on the logical resources available:

> At the time, there did not exist a generally recognized "mathematical logic" to which I could appeal, any more than it does today, where every . . . researcher has his own system.[16]

So Zermelo proposed to give the notion of *definit* an axiomatic treatment. The result was essentially second-order, as indicated by the clause that, if $P(g)$ is *definit* for each propositional function g (with only individual variables), then so are $\forall f(P(f))$ and $\exists f(P(f))$.

In his almost immediate response, Skolem[17] noted that Zermelo's new formulation was quite similar to his own (of 1922), the only major difference being the (uncritical) use of 'propositional functions'. Skolem wondered whether Zermelo intended to provide still more axioms to characterize those. Presumably, Zermelo did not think that a further formulation was needed—this is the material of which foundations are made.

Of course, Zermelo could have provided axioms to characterize the propositional functions. Indeed, around that time, Hilbert and Ackermann's *Grundzüge der theoritischen Logik* did this in their treatment of higher-order logic. But Skolem stands ready to interpret anything Zermelo might say as thoroughly first-order and thus subject to reinterpretation via the Löwenheim-Skolem theorem. The dialogue ended with this standoff.

[15] "Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre," in *Matematikerkongressen i Helsingfors den 4–7 Juli 1922* (Helsinki: Akademiska Bokhandeln), pp. 217–232; translated in Van Heijenoort, *op. cit.*, pp. 291–301). For an extensive discussion of Zermelo's axiomatization, see Moore, *Zermelo's Axiom of Choice: Its Origins, Development, and Influence* (New York: Springer, 1982).

[16] "Über den Begriff der Defintheit in der Axiomatik," *Fundamenta Mathematicae*, XIV (1929): 339–344. The quoted passage is from p. 340. See Moore, "The Emergence of First-order Logic," *op. cit.* Of course, today there *is* a "generally recognized" mathematical logic, the classical first-order predicate calculus. The irony is that Zermelo would not "appeal" to it, because he thought that it is too weak. In elaborating his view, Zermelo proposed a massive infinitary language.

[17] "Einige Bemerkungen zu der Abhandlung von E. Zermelo: 'Über die Definitheit in der Axiomatik'," *Fundamenta Mathematicae*, XV (1930): 337–341.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Skolem's "Über die mathematische Logik,"[18] an expository article, contains a few remarks on second-order logic. After developing a first-order language, he showed how variables ranging over "propositional functions" could be introduced; and he raised the possibility of quantifiers over these second-order variables:

> If "all" and "there exists" are applied to variable propositional functions, the question arises: what is the totality of all propositional functions (*ibid.*).

The latter question, of course, is the sort of thing Zermelo (and Second) did not think had to be raised. Once a domain is fixed, the range of the locution 'all subsets' is determined. After all, similar questions are not raised about the first-order variables and the other logical terminology, such as the connectives. Moreover, from this perspective, the only way to answer such questions is to *use* the corresponding terminology in the metalanguage. We say, for example, that $A\&B$ is true if A is true and B is true.

But for Skolem (and First) the question is legitimate and pressing. He asserted that only two conceptions for the range of the higher-order variables are "scientifically tenable." The first is, in effect, ramified type theory, in which propositional functions are associated with formulas of an expanding language. The second is to introduce the notion of propositional function axiomatically:

> The axioms will then become first-order propositions, since the . . . "propositional functions" . . . will assume the role of individuals. The relation between arguments and functions will then appear . . . as primitive (*ibid.*).

The debate is over the material of which foundations are made.

Of course, neither propositional functions nor sets are the first controversial entities to be studied by mathematicians. Negative, irrational, and complex numbers readily come to mind.[19] There are, to speak (very) roughly, three different stances that have been taken toward such entities by their proponents. Here, however, each has a troubling circularity. The first is simply to *postulate* the existence of the entities. Any axioms that are given are taken to *describe* the postulated entities. One insists on a distinction between the axioms and the entities themselves, and, thus, alternate interpretations are regarded as irrelevant. It need hardly be mentioned that postulation,

[18] *Norsk matematisk tidsskrift*, x (1928): 125–142; translated in Van Heijenoort, *op. cit.*, pp. 508–524.

[19] See E. Nagel, "Impossible Numbers: A Chapter in the History of Modern Logic," in *Teleology Revisited and Other Essays in the Philosophy and History of Science* (New York: Columbia, 1979), pp. 166–194.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

 Digitized by Arya Samaj Foundation Chennai and eGangotri

by itself, is not going to convince the wary. It begs the question if anything does. Often, postulation is accompanied by arguments concerning the usefulness of the entities, or the fruitfulness of the resulting theory. Much of Zermelo's and Cantor's writings fit this mold. The fruitfulness would apply to *any* structure that satisfies the axioms, however. Thus, the second possibility is *implicit definition.* One first gives axioms and then holds that the defined entities "are" or "can be" anything that satisfies them. In arithmetic, for example, Richard Dedekind's celebrated "The Nature and Meaning of Numbers"[20] defines a "simply infinite system" to be any collection of objects that has an operation with certain properties. He then gave a categoricity proof and defined 'the natural numbers' to be one such system. Skolem's second "scientifically tenable" attitude toward propositional functions, the axiomatic treatment, is also an implicit definition. The problem here is that the very issue at hand concerns just which systems satisfy which axioms. Sets, properties, propositional functions, or the like, are central items in that very enterprise. If the axioms used in the definitions are taken to be first-order, as Skolem explicitly intended, then many (nonisomorphic) systems are so defined, and there is no way to distinguish a preferred interpretation. This is relativism, and its confirmation. If the axioms are second-order, then the implicit definitions are circular. One "defines" the entities in question by using those same entities. Again, the understanding of the (object language) first-order variables is tied to the items in the logic. The third outlook toward problematic entities is *construction.* One shows how the objects in question can be taken as combinations of less problematic entities. Examples include the Dedekind construction of real numbers as sets of rationals, and the "definition" of complex numbers as pairs of reals. Skolem's first conception of propositional functions along the lines of ramified type theory is of this form. A series of languages is constructed and each propositional function is associated with formulas thereof. From the perspective of the advocates of higher-order logic, this construction is inadequate. Only countably many propositional functions are constructed this way. More importantly, virtually every construction appeals to the intuitive notion of set and, as Skolem acknowledged, natural number.[21] Again, these are the "problematic" entities under study.

[20] In *Essays on the Theory of Numbers,* W. W. Berman, ed. (New York: Dover, 1963), pp. 31–115 (originally published in 1888). See P. Kitcher, "Frege, Dedekind, and the Philosophy of Mathematics," in *Frege Synthesized,* L. Haaparanta and J. Hintikka, eds. (Dordrecht: Reidel, 1986), pp. 299–343.

[21] Recall that set theory is supposed to be the theory in which all of mathematics —including arithmetic—is (re)formulated. It would thus violate the envisioned

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

III

In philosophy, regresses like the ones encountered here are familiar. One cannot forever continue to define one's terms using other terms; one cannot forever continue to prove one's premises from still more premises; and, in the cases at hand, one cannot forever continue to codify the meaning of one's discourse by providing more metasemantics. The hope of foundationalism was that such regresses would end at the bedrock of self-evidence, but this was not to be. In the present situation, it looks like questions must be begged. The first-order camp rejects or, at any rate, challenges the very framework that the higher-order advocates use in explicating the meaning of informal mathematical discourse.

Several options are available at this point. One is to maintain that most of the discourse of informal mathematics can be taken at face value (i.e., not re-interpreted), and it successfully refers to structures that are unique (up to isomorphism). Such structures include those of the natural numbers, the real and complex numbers, and, perhaps, the set-theoretic hierarchy. I call this outlook *neutral realism,* because no position is taken concerning the nature of the indicated structures. Questions can be raised at this point concerning how the structures are *apprehended* or understood by individual mathematicians, how they are *characterized* or described in practice, and how descriptions and inferences about them are *communicated.* That is, how do we know that, say, all number theorists are talking about the same structures? Of course, it is not clear how far one can go in answering these questions while remaining philosophically neutral. But it is a constraint on any view called "realism" that such answers be possible.

My "Second-order Logic and Mathematical Practice" assumes neutral realism and argues from that perspective that first-order languages are inadequate. If the language of informal mathematics is

---

"foundation" to appeal to natural numbers in the formulation of set theory. Some of the early attempts to codify Zermelo's notion of *definit* function, such as those of H. Weyl ["Über die Definitionen der mathematischen Grundbegriffe," *Mathematisch-naturwissenschaftliche Blätter,* VII (1910): 93–95, 109–113); and Fraenkel ("Zu den Grundlagen der Mengenlehre," *Jahresbericht der Deut. Math.-Verein.* XXXI (1922): 101–2], used phrases like 'finitely many' which, of course, appeal to natural numbers. This was taken to be problematic. Zermelo ("Über den Begriff der Defintheit in der Axiomatik") claimed that his second-order formulation was free of such appeals. In the aforementioned expository article ("Über die mathematische Logik"), Skolem conceded that both of the "scientifically tenable" conceptions of propositional functions rely on numbers. He suggested that neither sets nor numbers be regarded as prior, but that the foundations for both be laid simultaneously "in an interrelated way."

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

understood as an uninterpreted first-order system, then the characterization and communication of the presumed structures is impossible. One "solution" is to maintain that the underlying languages of informal discourse are (at least) second-order, with standard semantics. This is to assume or presuppose that the second-order terminology is itself understood unambiguously, and that this is no more problematic than neutral realism in informal mathematics. As we have seen, Church held a view like this (and so do I). The presuppositions of second-order logic are those of classical mathematics, taken at face value.

Another option, still within the rubric of neutral realism, is to hold that mathematical structures are successfully characterized (up to isomorphism) and communicated, but to maintain that this occurs only intuitively, or only in informal practice. That is, the communication of mathematics is inherently informal.[22] This view thus denies a significant role to any formal semantics in explicating mathematical description and communication. One may then hold that, to the limited extent that mathematics can be codified, the underlying logic is first-order. Against Church, this combination rejects the connection between the presuppositions of classical mathematics, as practiced, and second-order logic. The former is accepted; the latter is not.

A third option is to reject (neutral) realism, and to embrace a version of Skolemite relativism. One maintains that mathematical theories do not have unequivocal interpretations—or even interpretations unique up to isomorphism. With a moderate amount of set theory, of course, one can characterize models of various theories, and, in some cases, it is possible to use the resources of set theory to specify an "intended" interpretation for a theory. Thus, one shows that the standard model of arithmetic is isomorphic to the set of finite ordinals. But, and here is the rub, this view invariably insists that the mathematical theories and metatheories have first-order semantics. Thus, any theory at any level, if satisfiable in an infinite domain, has models of every infinite cardinality, and ultimately there is no principled way to identify an intended one. In other words, the underlying model theory is yet another theory with lots of "interpretations" (presumably in its own metatheory). There

[22] See John Myhill, "On the Ontological Significance of the Löwenheim-Skolem Theorem," [*Academic Freedom, Logic and Religion*, M. White, ed. (Philadelphia: American Philosophical Society, 1951), pp. 57–70; also in *Contemporary Readings in Logical Theory*, I. Copi and J. Gould, eds. (New York: MacMillan, 1967), pp. 40–54] for a lucid presentation of this option. Similar views have been suggested (in correspondence or conversation) by several prominent logicians.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

is no unequivocal understanding for such terms as 'set', 'subset', 'finite', and quantifiers like 'all models'.

Paul Benacerraf[23] shows that Skolem shifted from the "mathematics is inherently informal" view, which he held in 1922 (*op. cit.*), to such a relativism. Another illustration of the dilemma is J. von Neumann's "Eine Axiomatisierung der Mengenlehre."[24] In the concluding section, he noted that it is easy to prove the categoricity of, say, Euclidean geometry, but von Neumann adds a parenthetical remark that every version of the theorem makes use of set-theoretic or, in present terms, higher-order terminology.[25] The theorems thus presuppose that such terminology is itself unequivocal and, perhaps, unquestioned. In examining this presupposition, von Neumann considers the possibility of a categorical set theory. One attempt at this was Abraham Fraenkel's[26] "axiom of restriction" that states, in effect, that no model of the theory is to have a proper subclass that is a model of the (other) axioms. Von Neumann objected that this axiom uses "the notions of naive set theory" and, thus, does not belong in the formulation of set theory itself—especially if it is to provide assurance that the antinomies have been avoided. To overcome this, a noncircular codification of the axiom of restriction must be provided. He proposed, first, that a "higher" set theory be developed axiomatically, one that contains the sets and proper classes of ordinary set theory as elements.[27] One can then use the relation of

[23] "Skolem and the Skeptic," *Proceedings of the Aristotelian Society, Supplementary Volume*, LIX (1985): 85–115.

[24] In *Journal für die reine und angewandte Mathematik*, CLIV (1925): 219–244; translated in van Heijenoort, *op. cit.*, pp. 393–413.

[25] Categorical axiomatizations of geometry include an axiom of continuity. One version is that every bounded, infinite *set* of points has a cluster point. Another is Hilbert's axiom of "completeness" that no *model* of the theory is to be a proper subclass of another structure that satisfies the (other) axioms [Hilbert, *Les principes fondamentaux de la géometrie* (Paris: Gauthier-Villars, 1982); see Moore, "The Emergence of First-order Logic"].

[26] "Über die Zermelosche Begründung der Mengenlehre," *Jahresbericht der Deut. Math.-Verein*, XXX (1921): 97–8. If suitably interpreted, the axiom of restriction "rules out" urelements, non-well-founded sets, and inaccessible cardinals. Notice the similarity between this axiom and Hilbert's axiom of completeness in geometry (*op. cit.*). Carnap and Bachmann's "Über Extremalaxiome" [*Erkenntnis*, VI (1936): 166–188; translated by H. G. Bohnert, *History and Philosophy of Logic*, II (1981): 67–85] is an insightful discussion of such axioms [see also Fraenkel, Bar-Hillel, and Levy, *Foundations of Set Theory* (Amsterdam: North-Holland, 1973, 2nd rev. ed.), pp. 113–9].

[27] A model of von Neumann's set theory would be a collection of sets (i.e., a class) together with the collection of its subclasses. But (proper) classes are not supposed to be elements of collections. Thus, talk of models requires a more encompassing system that has collections whose elements are the classes of ordinary set theory.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

"subset in the higher theory" to formulate the notion of a class satisfying various statements. He notes that the intersection (within the higher system) of all models of the (other) axioms might turn out to be itself a model of the axioms. If so, it satisfies the axiom of restriction, and no other model does. It is emphasized, however, that this construction is relative to the choice of the "higher" set theory and the system satisfying it. Other (perhaps still "higher") systems may have different classes of "models of the axioms" and, thus, different intersections and different models that "satisfy" the axiom of restriction.

This, of course, leads to the sort of regress or relativism that we encountered above, and there seems to be no (non-question-begging) reason to think it stops. Von Neumann concludes that

> . . . no categorical axiomatization of set theory seems to exist at all; for probably no axiomatization will be able to avoid the difficulties connected with the axiom of restriction and the "higher" systems. And since there is no axiom system for mathematics, geometry, and so forth that does not presuppose set theory, there probably cannot be any categorically axiomatized infinite systems at all (*op. cit.*, p. 412).

It is then noted that a similar relativism seems to apply to basic notions of cardinality, since even these concepts are formulated with variables which occur "with reference to the entire [set-theoretic] system." An example is the concept of "Dedekind-finite," which refers to "all subsets" of a given set. If this relativity is sustained, of the notion of finitude,

> . . . nothing but the shell of its formal characterization would . . . remain . . . It is difficult to say whether this would militate more strongly against its intuitive character, . . . or its foundation as given by set theory (*op. cit.*, p. 413).

The dilemma that von Neumann registers here corresponds to two parts of the present trilemma: (1) one can reject the assertion that the fundamental notions of number, cardinality, subset, etc., as formulated in practice, are unequivocal. The best grasp we have of these concepts is that codified in axiomatic set theory, taken to be uninterpreted. This is relativism. (2) One can maintain that we have an intuitive, unequivocal grasp of the notions and reject the adequacy of the axiomatic, set-theoretic foundation. I add another option: (3) one can maintain the intuitive character of the notion of subset and *use* this to provide second-order characterizations.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

IV

In practice, of course, the present regresses do not go on very long. No one, for example, has bothered to formulate a meta-meta-meta-language, and then wonder how to understand, or interpret it. The regresses end, on all accounts, in the discourse of informal mathematics. At this level, there is no further formal codification, perhaps because it is realized that this would only push the problems further back. It is not that we cannot precisely formulate the next metalanguage, but we see little point in doing so.

There is also almost universal agreement among present-day mathematicians concerning the practice of informal mathematics and the use of informal discourse. With the exception of traditional constructivists, there is no widespread and systematic disagreement over examples of correct proof, at least not now. There is at most an occasional skirmish. The more philosophical disputes noted here do not concern the correctness of informal mathematics, but rather things like how the discourse should be described, what it means, what it refers to, and what its nonlogical terminology is. None of the real or fictional disputants here should be regarded as advocating the revision of current practice.

This explains why the proof theories of the logics under examination here are remarkably similar, and underscores my opening remark that the differences between first-order logic and higher-order logic lie primarily in the different views on the totality of the range of the extra variables—in the "semantics." Indeed, one cannot prove more theorems using a second-order logic than one can with a first-order logic with set variables (and appropriate axioms). Both have admirably succeeded in "saving the appearances" concerning the correct use of informal discourse.[28]

Thus, to borrow a Quinean phrase, the regress ends in practice when we "lapse into the mother tongue" of informal mathematical discourse, and there things proceed rather smoothly, at least for now. This observation is rather congenial with the view held by some

[28] The deductive systems of common second-order theories are a bit stronger than their first-order counterparts. For example, one can prove the consistency of first-order arithmetic in second-order arithmetic. The reason is that the comprehension axiom scheme of second-order logic has instances that contain bound predicate variables. These impredicative definitions yield classes that cannot be defined in the first-order languages and, thus, do not fall under the first-order axiom schemes. Second-order arithmetic is equivalent to first-order analysis, however, and first-order set theory is much stronger than second-order analysis. On the comparison of first-order set theory with second-order set theory, see my "Second-order Languages and Mathematical Practice" and *Foundations without Foundationalism*.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

philosophers that there is no more to understanding concepts than knowing how to use the relevant language. Variations on this theme have been championed by Hilary Putnam, Michael Dummett, and the later Wittgenstein.[29]

A slogan often associated with such views is "meaning is use," but I believe this is misleading. For present purposes, it is not a thesis about meaning, but about understanding. The claim is that understanding should not be ineffable. One understands the concepts embodied in a language to the extent that she knows how to use the language correctly. Call this the *use thesis*.

In the philosophy of language, the most straightforward opponents of the use thesis are those who tie understanding to a grasp of something that transcends use, or is conceptually independent of use. One example, perhaps, is Gottlob Frege, who held that each meaningful expression of an interpreted language is associated with a timeless, eternal, objective, and mind-independent entity called its *sense*. The concepts associated with a part of language are understood only if the requisite senses are grasped. To be sure, a given person's ability to use a language, such as that of arithmetic, is evidence for his having grasped the senses, but use and understanding are two different things. A similar view seems to underlie Noam Chomsky's distinction between competence and performance, but a consideration of this would take us too far afield. Against such views, the use thesis is that the ability to use a language constitutes understanding and thus grasping the concepts.

As Dummett puts it:

> An individual cannot communicate what he cannot be observed to communicate: if an individual associated with a mathematical symbol or formula some mental content, where the association did not lie in the use he made of the symbol or formula, then he could not convey that content by means of the symbol or formula, for his audience would be unaware of the association and would have no means of becoming aware of it (*op. cit.*, p. 216).

Moreover, in learning a language, such as that of arithmetic, one does not directly learn the sense of each expression, much less the denotation of each term. One learns, in effect, how to use each expression: how to perform the computations that underlie simple

[29] Putnam, "Models and Reality," *Journal of Symbolic Logic*, XLV (1980): 464–482); Dummett, "The Philosophical Basis of Intuitionistic Logic," in *Truth and Other Enigmas* (Cambridge: Harvard, 1978), pp. 215–247; Wittgenstein, *Remarks on the Foundations of Mathematics* and *Philosophical Investigations*, G.E.M. Anscombe, trans. (New York: MacMillan, 1958).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

equations, how to apply arithmetic to everyday problems, and, later, how to prove and refute arithmetic statements and reason hypothetically in the language:

> These things are all that we learn when we are learning the meanings of the expressions of a language of the mathematical theory in question, because they are all that we can be shown (*op. cit.*, p. 217).

Perhaps another way to characterize the present contrast is to note that, on Fregean views, it is possible, at least in principle, for someone (or some machine) to be able to use a language correctly and still not grasp the relevant concepts. One can "fake it." Dummett writes:

> To suppose that there is an ingredient of meaning which transcends the use that is made of that which carries the meaning is to suppose that someone might have learned all that is directly taught when the language of a mathematical theory is taught to him, and might then behave in every way like someone who understood the language, and yet not actually understand it, or understand it only incorrectly (*op. cit.*, pp. 217–8).

Conversely, on the views under attack, it seems to be possible to grasp concepts and yet not be able to say anything correct about them. Again, understanding is conceptually independent of use, even if the gap is usually not very large in practice. The use thesis rejects both of these possibilities out of hand.[30]

It is often thought that the use thesis stands opposed to realism, the view that the variables of the language range over a realm of objects that exist independent of the mathematician and her mental life. Indeed, the slogan, "meaning is use," is often taken as a hallmark of antirealism. Dummett, for example, argues that the use thesis suggests, or even demands, that Tarskian semantics, with domain of discourse and truth conditions, is inappropriate. He proposes that 'use' replace 'truth' or 'satisfaction' as the main ingredient of a compositional semantics. One provides "proof conditions" instead of "truth conditions."

[30] Admittedly, these possibilities-in-principle invoke a rather broad metaphysical modality. It is open to a Fregean to defend a thesis that a complete fake, or even a large fake, is impossible in some, perhaps epistemic, sense. Plato may have held such a view in that the Socratic method seems to presuppose that, if a person fails to grasp a concept, then there are situations in which he cannot give intelligent responses. As for the converse, it is surely reasonable to hold—again on epistemic grounds—that there is a limit to the gap between competence and performance. It might be noted that the use thesis can itself countenance a distinction between competence and performance, but both are related to a subject's ability to use the language. Competence refers to global use and the ability to relate the relevant aspects of language to the rest of it. Performance concerns particular utterances in restricted areas, which may or may not be consistent with global use. On this sketch, the distinction between competence and performance is one of degree, not of kind.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

I do not believe that this broad conclusion is warranted. The use thesis is quite plausible, and places sensible constraints on any theory of meaning. But, by itself, it does not demand an antirealist ontology, nor a non-Tarskian semantics. It does not follow that 'use' must be the central item in semantics. It does follow that ontology and truth conditions must be compatible with the learnability, and understanding of a language through its use. Truth conditions should not be that far removed from the knowledge implicit in the correct use of a language, and from whatever it is that one learns when acquiring facility with a language. There should be a natural (if not inevitable) link between use and truth conditions. Ideally, it should be clear, or at least defensible, that an object-language theory, plus its semantics, is a further articulation of the object-language theory alone, not an abrupt shift away from it. To use a Wittgensteinian metaphor (probably out of place), it should be plausible that, in adding semantics to a theory, one has "gone on as before." Without this, Dummett is correct that meaning is ineffable, or at least the charge is left unrefuted.[31]

In a similar context, Putnam notes that an advocate of the use thesis does not have to reject notions like model and reference. It is a question of understanding the roles these play in philosophy of mathematics. The grasp of particular models cannot be the central criterion of understanding. The model theory is itself formulated (originally) in informal language, which is understood through its use. Once one

> . . . has succeeded in understanding a rich enough language to serve as a meta-language for some theory T, he can define 'true in T' . . . he can talk about "models" for T, etc. He can even define 'reference' . . . exactly as Tarski did (*op. cit.*, p. 479).

The model theory and the formal semantics extend the resources of the original language, and may even articulate it further, but both come after the language is understood, and the original understanding does not consist of grasping a model, intended or otherwise. Understanding the language—original and extended—is knowing how to use it.

[31] In my "Logic, Ontology, Mathematical Practice" [*Synthese*, LXXIX (1989): 13–50], I argue that structuralism is compatible with realism, Tarskian semantics, *and* the use thesis. See essay 3 of Parsons' *Mathematics in Philosophy* for an insightful discussion of the sense in which the arithmetic within the theory of arithmetic-plus-arithmetic-truth is the same or different from arithmetic *simpliciter*.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

At the level of informal discourse, prior to semantics, one can ask about the "meanings" of certain words and their referents. Many of the answers turn out to be platitudes like '12 is a natural number' or '6 is the sixth natural number after zero'. Similarly, the informal metalanguage is capable of distinguishing standard from nonstandard models (or semantics): "Standard models of arithmetic are those which are isomorphic to the natural numbers. Nonstandard models are those which begin with a copy of the natural numbers, followed by other elements of the domain." We actually say things like this to students of logic, and some of them understand it, or seem to. From the perspective of the regress above, of course, it begs the question (like everything else). The "problems" concerning meaning and reference apply to those very statements. We cannot rule out unintended interpretations of the metalanguage, unless we can somehow fix its semantics, and we cannot do that without using more language, in which case we must worry about its interpretation. And on it goes. But the use thesis blocks this regress or, at any rate, decides to stop by refusing to carry it further. We use the languages without (further) interpretation, but, to echo Wittgenstein, it does not follow that we use them without right. Both (informal) metalanguage and object language are understood when we know how to use them. We make the statements, we understand them, and we mean what we say. It makes no sense to ask for the meaning and reference of the entire informal language and to expect the answers to be somehow independent of that language.

Recall that two of the options in the above trilemma involve what I call *neutral realism,* the view that mathematical discourse is to be taken at face value. Arithmetic is about natural numbers, analysis is about real numbers, set theory is about sets, etc. This much is compatible with the use thesis and, indeed, seems to follow from it. What else is the use theorist going to hold? To be explicit about it, the informal, or preformal language of mathematics has a predicate N for 'is a natural number'. It is a basic item of the lexicon. The mathematician says that four is a natural number, $N4$. It follows, by existential generalization, that $\exists\, xNx$, numbers exist. Period. I propose that, on the use thesis, Skolemite relativism is to be rejected out of hand. The possibility of reinterpretation of the language has nothing to do with original use and, thus, nothing to do with understanding. The informal language, as it is used, is not equivocal.

At some point, it seems, we asked for a formal semantics to be used in accounting for how structures are apprehended, characterized, and communicated. As Putnam notes, this is the fatal step. It is not that the desire for a formal semantics is somehow illegitimate. Quite

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

 Digitized by Arya Samaj Foundation Chennai and eGangotri

the contrary. The problem is with the idea that formal semantics is needed, or even helpful, in giving an account of what we grasp when we learn the meaning of a language:

> To adopt a theory of meaning according to which a language whose whole use is specified still lacks something—viz. its "interpretation"—is to accept a problem which *can* only have crazy solutions. To speak as if *this* were my problem, "I know how to use my language, but, now, how should I single out an interpretation?" is to speak nonsense (*op. cit.*, pp. 481–2).

If this conception of understanding is to be sustained, there is an interesting problem analogous to that of reference and nonstandard models. Use theorists are often criticized for leaving the concept of "use" vague. Surely an account is needed, even if problems of interpreting that account inevitably arise. As Wittgenstein[32] puts it:

> It all depends [on] *what* settles the sense of a proposition . . . The use of the signs must settle it; but what do we count as the use? (*ibid.*, pp. 366–7)

It might appear that attention is shifting from semantics to proof theory, but this is misleading. The problem is to account for how the understanding of a language of mathematics, through its use, constrains the future use of the language. How do we "go on as before"? Specifically, what is the relationship between formal proof and broad, global "use," as it figures in understanding? I maintain that the use thesis and formalism make a very bad philosophical partnership.

There is an important aspect of mathematical practice—the *use* of mathematical discourse—that is not captured by first-order languages, namely, the sense in which the discourse outstrips, or even "transcends," its previous manifestations. I would also suggest that many of the considerations in favor of second-order languages can be straightforwardly understood in these terms.[33] The arguments that certain concepts and theories resist first-order treatment amount to claims that first-order languages are not adequate to capture important aspects of the use of the relevant languages. I take up an example presently. To anticipate the conclusion, even if the use thesis dissolves the above regress of semantics, it does not dissolve the issue of second-order logic. The problem of interpreting the interpretations is replaced with the problem of rules for following rules.

[32] *Remarks on the Foundations of Mathematics.*
[33] See my "Second-order Logic and Mathematical Practice" and *Foundations without Foundationalism.*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

There is an interesting affinity between the way the problem of second-order languages is posed above and the discussion of rule following in Crispin Wright's *Wittgenstein on the Foundations of Mathematics.*[34] In the above scenarios, the two characters Second and First seemed (at times) to be in agreement. They accepted the same verbal description of the interpretation of the language. But they later discovered that they were interpreting the description (of the interpretation) differently. Thus begins the regress. Wright presents a thought experiment involving two people who agree on a description of a rule, and yet go on to apply it differently. There is a hidden divergence on the interpretation of the description of the rule. Let us pursue the analogy.

Saul Kripke[37] formulates an interesting problem on Wittgenstein's behalf. Most of us were taught how to add in grade school, and we now employ the symbol '+'. The problem is to account for our meaning the addition function by '+'. The thesis at hand is that the meaning of the term is fixed by our use of it and other expressions. But "what do we count as the use?" The addition function has infinitely many arguments and values. We have not "used" all of them yet, nor will we ever. Let 'quus' denote a binary function that agrees with addition for all arguments less than some large number N, but differs thereafter. I suggest that the role of the quus function here is analogous to the role of nonstandard models in the above discussion. Kripke's version of the problem is to show how we can refute a "skeptic" who tries to claim (ad hominem) that, in the past, we really meant quus by '+'. The analogy is with a skeptic (i.e., Skolem) who tries to claim that our past use of mathematical discourse may have referred to a nonstandard model.

If the value of N is large enough, both the plus function and the quus function are consistent with all of our past activities—all the numbers we have added (or quadded) so far. Both are also consistent with our physical dispositions to add (assuming that there is some limit to the size of numbers we are disposed to consider). Moreover, anything we have said about our practice, such as the recursive definition of '+', is subject to alternate interpretations, supporting either the "plus" or the "quus" reading.

Putting aside the exegetical and philosophical merits of Kripke's proposed (antirealist) solution, the problem points to a difficulty for the use thesis, or at any rate, a presupposition behind the use of ordinary language. Virtually every (nonfictitious) person believes that we do mean plus, and not quus, by '+'. It follows that our ability

[34] (Cambridge: Harvard, 1980), ch. 2. I owe this observation to Barbara Scholz.
[35] *Wittgenstein on Rules and Private Language* (Cambridge: Harvard, 1982).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

 Digitized by Arya Samaj Foundation Chennai and eGangotri

to add, and our knowledge of arithmetic, somehow go beyond our previous performance, our dispositions to behave, etc., or at least we assume that they do. When we encounter new instances of addition and handle them correctly, we have "gone on as before." Wittgenstein's problem is to maintain the use thesis and yet have the pattern of "use" of an expression at a given time go beyond its previous manifestations. We need a plausible account of what it is to "go on as before." I think this is the meaning of the first passage quoted at the top of this paper. The problem is to maintain realism in the sense that we do mean something by our terms—Skolemite relativism is surely absurd—and yet not be limited by the "observation" of our previous behavior, not "empiricism."

The same sort of difficulty arises in the present situation, but now on a grand scale. Here it is "realism but not formalism, that is the hardest thing." Consider, for example, real analysis and, to make the analogy better, think of the completeness axiom as sanctioning a rule:

> When one learns that a set S of real numbers is bounded from above, one can infer that S has a least upper bound.

Suppose that some mathematician, or some community of mathematicians, formulates analysis, learns the completeness rule, and uses it successfully for several decades. Call him (or them) "Karl." Suppose also that, during those decades, analysis is the most powerful theory Karl has. In particular, suppose that Zermelo-Fraenkel set theory has not been developed.

Now, a first-order formulation of real analysis takes the completeness rule as a scheme, one instance for each open formula of *Karl's language*. Like Kripke's quus function, first-order analysis characterizes (1) Karl's past (correct) use of the completeness rule and (2) his physical dispositions to apply the rule (correctly). But it captures much more: (3) every use of the rule he *can* make given his expressive resources: here the modal term 'can' is not limited to physical possibility—the first-order scheme includes every application of the rule in Karl's present language; (4) his ability to describe his possible uses; (5) our ability to describe his ability to apply the rule (correctly); etc.

It should come as no surprise, then, that, from Karl's perspective, there is no significant difference between the class of theorems he can prove in first-order analysis and the class he can prove in second-order analysis. The first-order scheme has instances covering just about any use he can make given his (or their) resources. That is to say, any use he can make *at this stage* in his development. Yet, as

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

with Kripke's quus function, this is not enough. The problem is to maintain the use thesis and yet hold that the "use" of at least some expressions (like the completeness rule) goes beyond, say, (1) to (5) above.

Suppose that Karl now develops, or becomes aware of, new expressive resources. For example, he may formulate some set theory and model the real numbers therein. Suppose also that he uses his extended language to define sets of real numbers he could not define prior to the new developments. He then applies the completeness rule to these sets and thereby learns more analysis—he learns more about the real numbers. It seems evident, at least to me, that he (or they) has "gone on as before." Karl's reaction to the newly defined sets of real numbers is exactly analogous to ours when we encounter a particular sum for the first time. We apply the same rule to a new case.

Similarly, Parsons' forthcoming "The Uniqueness of the Natural Numbers" describes a scenario in which two (communities of) mathematicians meet and learn each other's language. Each proves that the other's natural and real numbers are isomorphic to his own. This is not accomplished by an esoteric appeal to intuition. Rather, each uses the resources of the combined language to define sets of numbers, and then applies the relevant axiom (of induction or completeness) to these sets. That is, each mathematician applies the relevant "rule" to cases he could not formulate before the meeting. As with Karl and set theory, the combined language expands the resources, and provides new cases to which the rule can be applied, but the rule itself is not changed.

Second-order analysis captures this intuition of "going on as before." For Karl, the terminology of set theory is involved in extending the comprehension scheme (of second-order logic) that is used to describe and establish the existence of the "new" sets of real numbers. Once this is accomplished, the completeness rule is applied to these sets—the same completeness rule that Karl used before. The rule itself is not extended, and not reinterpreted.

With first-order analysis, the situation is not as straightforward. To derive the new theorems, the set-theoretic terminology must be used to produce new instances of the completeness scheme itself. That is, Karl must produce new cases of the very defining statements of analysis. In effect, the new *language* entails that Karl is working in a new *theory*, with a new set of rules. It is an extension of the old one, perhaps, but that theory has lots of extensions (since it is incomplete). What is lacking is a principled account of how the chosen new theory is the "same"—how Karl went on as before. It is straightfor-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ward to prove that there are models of the original first-order theory in which some of the new theorems are false. Thus, if Karl's (prior) use of the discourse of analysis involves no more than what is captured in the first-order theory (with its perfect match between semantics and proof theory), then the "new" theorems are not "about" the real numbers.

In sum, the first-order view and the second-order view seem to diverge in practice when the expressive resources are expanded. The first-order theory is limited to the sets of real numbers that are definable in the original language. The second-order account applies the same rule to the newly defined sets. Perhaps we did not note the potential divergence in the above scenarios because it was assumed that the total expressive resources are held constant throughout the regress, or maybe we silently allowed our advocate of first-order logic to reconstruct the basic theories at each level.[36]

I do not claim that adopting higher-order logic somehow solves the twin problems of rule following and Skolemite relativism. It does not refute the skeptics. If our character First were reintroduced at this point, she would argue that the second-order language creates a false appearance of "going on as before." With the adoption of the set theory, the completeness rule is the same, perhaps, but the comprehension scheme changes. She would claim that there is a shift in the range of the second-order variables. Recall that she denies that this range is fixed. It changes from one model of the semantics to another.

My thesis here is that, for better or worse (well, for better), the attitude underlying the practice of mathematics is that skepticism is false. We do go on as before, both in following rules and in extending our theories. At any rate, we talk and act that way. It may be an illusion, I suppose, but when philosophy is said and done, we do not take skepticism seriously. Second-order languages echo this presupposition. It is in line with Church's statement that the assumptions of second-order logic may be more conspicuous than those behind the practice of classical mathematics, but they are not more troublesome.

[36] There is an interesting analogy between these considerations and a major argument in Michael Friedman's *Foundations of Space-time Theories* (Princeton: University Press, 1983) for the reality of space-time. Friedman concedes that, if we focus attention exclusively on the most powerful and complete physical theory we have, general relativity for example, the substantival versions that refer to space-time are equivalent to the relationalist versions that do not. But space-time is needed to see the smooth relationships between, say, gravitational theory alone and gravitational theory plus electro-dynamics. The relationalist theories do not go together that well, and the reality of space-time is confirmed in the historical advance from the weaker theory to its more comprehensive successor.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Similar considerations rule out a formalistic account of the "use" of mathematical discourse. No single explicitly formulated algorithm can capture Karl's prior understanding and use of his language. If nothing else, the incompleteness theorem establishes this. No algorithm can account for Karl's "going on as before" when his resources are expanded. At best, an explicit algorithm can codify the applications of Karl's theories in their present state of articulation —in their present languages.

Benacerraf puts the situation well:

> We *do* need a metaphysically and epistemologically satisfactory account of the way mathematical practice determines or embodies the meaning of mathematical language. (We could also use a satisfactory account for areas other than mathematics.) We may even need to devise new concepts of meaning to forge such an account.
>
> . . . mathematical practice reflects our intuitions and controls our use of mathematical language in ways of which we may not be aware at any given moment, but which transcend what we have explicitly set down in any given *account*—or may ever be able to set down.
>
> With Gödel, I incline toward this view. But I am sufficiently aware of its vagueness and inadequacy not to be tempted into thinking it constitutes a *view*. It is merely a direction (*op. cit.*, pp. 110–1).

I also incline to this "view" and, as far as I have put it here, I think the later Wittgenstein did as well. The difference with Gödelian Platonism (and with formalism) is the extent to which we are able to fix the use of our terms completely. What is needed here is a synthesis between "relativism," which holds that the constraints on the use of our language do not extend much beyond our present ability to describe and apply our rules, and traditional Platonism, which holds that our minds can somehow grasp infinite structures, and that the concepts are thereby fixed completely, once and for all. The question concerns the extent to which our present practice in, say, analysis, determines the future use of the terms or, to use another Wittgensteinian phrase, the extent to which meaning and use are and are not fixed "throughout logical space."

However this is resolved, I make the modest proposal that we can shed some light on the requisite concepts of meaning and reference, and the sense in which these can outstrip prior manifestations, by studying the use of second-order concepts in mathematics.

STEWART SHAPIRO
Ohio State University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

COMMENTS AND CRITICISM

SENSATIONALIZED PHILOSOPHY: A REPLY TO MARQUIS'S "WHY ABORTION IS IMMORAL"*

IN a recent article, Don Marquis[1] claims to show "Why Abortion is Immoral." The title is, as I shall show, much bolder than what is warranted by his argument. This essay simply rebuts Marquis's initial assumption: that the only important question for settling the abortion issue is the moral status of the fetus, and the corroboration for this claim which he alleges to exist in the abortion literature. I mean this as only a partial reply,[2] and I do not claim to add to the positive argument here. In fact, I would not add to the thousands of pages written on this topic if I did not feel morally compelled to rebut what I see as a sensationalized and erroneous treatment of an issue of utmost current political importance.

In the beginning of his article, Marquis explicitly makes the assumption that "whether or not abortion is morally permissible stands or falls on whether or not a fetus is the sort of being whose life it is seriously wrong to end" (183). There are two issues that arise in assessing this claim: (1) Is the fetus the sort of thing that could have any rights or toward which we could have obligations? (2) Are these rights or obligations prima facie or absolute? Marquis's claim seems to conflate the two issues, assuming that whatever rights of, or obligations to, fetuses there are, they must be absolute.

The assumption that any obligations we have to fetuses are absolute, presented without argument, is philosophically (and politically) irresponsible. It is as if fetuses were things growing out in the garden, and the question of abortion were whether one may decide to till them under rather than let them come to fruition. The question of abortion inextricably involves (at least) two lives and a compelling bundle of rights on the side of the woman carrying the fetus, whatever we decide about the status of the fetus. Ignoring these rights makes about as much sense as considering the issue of the moral permissibility of killing adult, fully-conscious humans without considering the justification of self-defense. If no countervailing rights

* I would like to thank Neal Becker, Jack Bricke, Tony Genova, Leslie Jones, and JoAnn Reckling for helpful comments. I, alone, am responsible for any remaining errors.

[1] This JOURNAL, LXXXVI, 4 (April 1989): 183–202.

[2] This is only part of what was originally a longer reply, which is available on request from the author.

0022-362X/90/8705/262–4 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

or other moral considerations were to be allowed to figure into the judgment, the criterion on which Marquis claims that abortion is impermissible, having a "future-like-ours," would also rule out killing in self-defense. He makes no exceptions to his claim that abortion is immoral. Thus, it seems that to be consistent he would also have to reject any self-defense plea in killing any human being.

The only justification Marquis provides for the dubious assumption that a fetus's right to its future is absolute and overriding is an appeal to his favorite authors on abortion. He writes:

> Many of the most insightful and careful writers on the ethics of abortion . . . believe that whether or not abortion is morally permissible *stands or falls* on whether or not the fetus is the sort of being whose life it is seriously wrong to end (183, emphasis mine).

The list of authors Marquis provides is quite impressive. The claim he makes about their arguments here is false, however. They all consider rights (and other moral concerns) of the pregnant woman, and find cases in which her rights or concerns override any considerations in favor of the fetus.[3] The authors cited can be divided into two categories: those who argue that abortions are permissible because the fetus is not the sort of being whose life it is seriously wrong to end,[4] and those who argue that abortions are impermissible in most (but not all) cases because the fetus is such a being.[5]

[3] Mary Anne Warren, "On the Moral and Legal Status of Abortion," *The Monist*, LVII, 1(1973): 43–61: "The immorality of abortion is no more demonstrated by the humanity of the fetus, in itself, than the immorality of killing in self-defense is demonstrated by the fact that the assailant is a human being" (p. 46). H. T. Engelhardt, Jr., "The Ontology of Abortion," *Ethics*, LXXXIV, 3 (1974): 217–234: "The rights of the mother regarding abortion are paramount. After all, she is the only actual person involved. . . . Consequently, it is to her that one owes overriding obligations" (p. 233). Michael Tooley, "Abortion and Infanticide," *Philosophy and Public Affairs*, II, 1 (1972): 37–65. Tooley claims that the argument is settled by the fact that the fetus is not a person, and endorses Judith Jarvis Thomson's general claims about what would be the case if the fetus were a person, namely, that the woman's rights override. Joel Feinberg, "Abortion," in *Matters of Life and Death: New Introductory Essays in Moral Philosophy*, Tom Regan, ed. (New York: Random, 1986), pp. 256–293: "Even if we grant that the fetus is a moral person and thus has a valid claim to life, it does not follow that abortion is always wrong" (p. 233). J. T. Noonan, Jr., *Private Choice* (New York: Free Press, 1979). Noonan urges the reversal of Roe v. Wade, but on many grounds, including the rights of the family and the poor, as well as his view that fetuses are the sorts of beings that it is seriously wrong to kill. Philip Devine, *The Ethics of Homicide* (Ithaca: Cornell, 1978). Devine allows that there are cases when abortion is not morally impermissible, see esp. ch. III. It also seems that one cannot legitimately ignore Thomson, "A Defense of Abortion," *Philosophy and Public Affairs*, I, 1 (1971): 47–66, as being among the most insightful writers on abortion. Her view is quite clearly opposed to Marquis's; the legal, if not in each case moral, permissibility of abortion stands on the rights of women to refuse to donate their bodies to others.

[4] Warren, Engelhardt, Tooley, and Feinberg.

[5] Noonan and Devine.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

With regard to the first group, it is an obvious logical mistake to infer from:

(A) Since the fetus is not the sort of being whose life it is seriously wrong to end, it is morally permissible to abort.

that therefore,

(B) If it were the case that the fetus is the sort of being whose life it is seriously wrong to end, then it would be morally impermissible to abort.

The writers in the first group hold (A), but they make no claims like (B). That is, these writers merely claim that the fetus's lack of personhood is a sufficient, though not necessary, condition for abortion to be morally permissible. Marquis's phrase 'stands or falls' requires it to be both a necessary and sufficient condition on the permissibility of abortion.

In the second group, both authors deny (A), but allow that, in cases in which the mother's life is in danger, or, perhaps, the woman was a victim of rape, abortion is morally permissible, thus denying (B) as well. In these cases, they reason, the woman's right to life or her serious loss of well-being overrides the rights of the fetus. Thus, they regard the woman's rights as a relevant issue in deciding the moral permissibility of abortion, and so could hardly be said to argue that the moral permissibility of abortion stands or falls with the issue of whether it is wrong to kill fetuses.

Even if we might ultimately agree that the fetus is the sort of thing whose killing is so morally wrong as to overwhelm completely a woman's rights to privacy, health, medical care, and even life, the point surely needs argument. Nothing that has been said in the abortion debate to date has come close to settling this issue against the woman. So *at most* Marquis can claim to have shown "Why Abortion is Killing a Being-Like-Us." When one recalls that persons may legitimately be killed for many reasons, this title has not the same moral urgency of Marquis's.

ANN E. CUDD

University of Kansas

DOES A FETUS ALREADY HAVE A FUTURE-LIKE-OURS?

Some of the most interesting and underexplored issues in philosophy are those of how human beings are in time. A person's relationship to her future is very complex, particularly if time

0022-362X/90/8705/264–8 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

passes, as we commonsensically believe that it does. In "Why Abortion Is Immoral," Don Marquis[1] argues that what makes killing a person wrong is that it deprives the person of her future. He concludes that abortion is wrong because it deprives the fetus of a "future-like-ours." The line of argument is clear.

> The future of a standard fetus includes a set of experiences, projects, activities, and such which are identical with the futures of adult human beings and are identical with the futures of young children. Since the reason that is sufficient to explain why it is wrong to kill human beings after the time of birth is a reason that also applies to fetuses, it follows that abortion is prima facie seriously morally wrong (192).

The unexamined premise in the argument is that a fetus *already* has a future-like-ours of which it can be deprived.[2] For the argument to be convincing, it is necessary that a fetus *at its time* "possess" or be related to a future-like-ours in a way that allows the transfer from the wrongness of killing us persons to the wrongness of killing fetuses.

Fetuses are very different from normal adult humans. The connections between a fetus at an earlier time and a person (or person stage) at a significantly later time are very different from the connections between the person stages at different times which compose one person. Philosophical investigations of personal identity through time have revealed the complexity of the biological and psychological connections between the earlier and later stages of one person. These significant differences invalidate the claim that a fetus has a personal future in the same way that a normal adult human has a personal future.

The differences between a person's relationship to her future and a fetus's relationship to its future are striking even when the passage of time is ignored. In B-series time (a time that is composed entirely of earlier and later temporal locations with their occupants), an earlier person stage has many relations with later person stages which make these later person stages be "her future." The most widely considered relations in contemporary discussions of personal identity are those of memory, continuity of character, and intention-to-action.[3] Memory relations are from later person stages to earlier person stages. The later person stages are able to remember the experiences of the earlier person stages or there is an overlapping

[1] This JOURNAL, LXXXVI, 4 (April 1989): 183–202.
[2] "Since a fetus possesses a property, the possession of which in adult human beings is sufficient to make killing an adult human being wrong, abortion is wrong" (202).
[3] See Derek Parfit, *Reasons and Persons* (New York: Oxford, 1984), pp. 205–7.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

chain of such memory connections (memory continuity). The relation of continuity of character is that in which later person stages either have a character similar to the earlier person stages or are different in ways that are explicable by the operation of normal causes.[4] The relation of intention-to-action is that between an earlier intention and a later action that carries out that intention. Normal adult humans have all sorts of plans and projects for their short- and long-term futures which take time to implement.

There are other relations that connect earlier and later person stages. Some of the "mental processing" that ordinarily goes on in normal humans, such as forming generalizations from repeated observations or "digesting" an emotionally charged experience, takes a significant amount of time and so can be considered to include relations between person stages. In addition, there are all of the neurophysiological relations that underlie the ordinary continuation of mental life in persons. That a person has pretty much the same beliefs, wants, skills, and habits that she had 30 minutes (or 30 days) earlier depends upon a similarity of neurophysiological conditions between the earlier and later person stages.

Most of these relations to later person stages exist even when the earlier person is asleep or temporarily unconscious. Even intentions to perform later actions might be considered to continue through periods of unconsciousness. Since a temporarily unconscious person is still strongly related to her future, to kill her while she is unconscious is to deprive her of her future.

Young infants do not have all of the psychological complexity that adult persons have. Nevertheless, young infants are commonsensically understood to have perceptions, beliefs, desires, and emotions (whether or not the experimental data confirms this) and to learn from experience. For this reason, the neurophysiological states and processes of young infants can be understood to underlie something like the ordinary continuation of mental life in persons. A good case can be made that young infants are related in some (though not all) ways to a personal future.

The situation of a fetus at an early stage of development is very different.[5] A fetus at an early stage of development has neither a mental life of feelings, beliefs, and desires nor a developed brain and nervous system. There are none of the main relations with a personal future which exist in persons. Although there is some biological continuity between them so that there is a sense in which the later

[4] See *Reasons and Persons*, p. 207, for a brief discussion of "normal causes."
[5] As the fetus develops, it becomes more similar to a young infant, and so progressively acquires more of a relationship to a personal future.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

person stages "are the future" of the fetus, the fetus is so little connected to the later personal life that it can not be deprived of that personal life. At its time the fetus does not already "possess" that future personal life in the way that a normal adult human already "possesses" his future personal life.

Our commonsense views about time and entities in time involve a past that is fixed and determinate, a future that includes alternative possibilities whose actualization may be affected by action, and a process of what is future becoming present and past. In a time of this sort, how an entity "has a future" is more complex because when some temporal part (stage) of the entity is present, there is not a fixed and fully determinate future to which that entity can now be related. With respect to persons, a present person stage is not now related to a specific determinate later person stage that will become present. A person's future includes a branching range of possibilities (including his death) from which only one life course can become present. The branching of possibilities is such that most outcomes can become present only if certain earlier possibilities have become present. The actualization of one possibility makes available those later possibilities which presuppose it. This is particularly pronounced for the acquisition of skills, abilities, and capacities, which open up new ranges of possibilities that would otherwise not be available.

Many factors external to the person affect which of the person's possibilities become present. A normal adult human has only limited control over which of his possibilities become present. This control which a person exercises and attempts to exercise over his future is the most important connection that now exists with a specifically personal future. This control also depends upon the person's wants, skills, abilities, and capacities. The person wants various things for his future (including wanting himself to act) and exercises his powers to affect what happens.

A fetus is separated from a personal future by many "layers" of possibility. The possibilities that are available to a person or even to a young infant are not now available to the fetus. Only if the fetus develops in the right ways (favorable possibilities become present) will it acquire the capacities that make available the infant's possibilities. A great deal of favorable development would be necessary before the fetus could control its future in the way that persons do. The fetus does not now have a personal future.

Marquis has succeeded in formulating an important feature of people's opposition to abortion: the notion that abortion "cuts off" the fetus's future. A close examination of what it is "to have a future" reveals that at its time a fetus does not have a personal future

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of which it can be deprived. A living human cell that might be stimulated to develop into a clone of a person does not now have a personal future. A fetus similarly has only the potentiality to develop a personal future. For this reason, killing a fetus is morally very different from killing a normal adult human.

PETER K. MCINERNEY

Oberlin College

KILLING, ABORTION, AND CONTRACEPTION: A REPLY TO MARQUIS*

Don Marquis, in "Why Abortion is Immoral,"[1] argues that "abortion is, except possibly in rare cases, seriously immoral, that it is in the same moral category as killing an innocent human being" (183). His argument for this is that abortions share with killings the central feature that makes them wrong. "Killing someone is wrong, primarily because the killing inflicts (one of) the greatest possible losses on the victim . . . the loss . . . of all those activities, projects, experiences, and enjoyment which would otherwise have constituted [the victim's] future personal life" (189). Marquis argues that, since fetuses have futures that include a set of "experiences, projects, activities and such which are identical with the futures of adult human beings," it follows that abortion is "prima facie seriously morally wrong" (192). Marquis answers several possible objections to his account. In this paper, I shall discuss the one objection to which his answer is clearly inadequate. I shall claim not only that his answer is inadequate, but that an adequate answer is not available to him, and thus that his account fails to support the claim that the overwhelming majority of abortions are seriously immoral.

I

Marquis admits that, if his account of the wrongness of killing and of abortion entails that contraception is also seriously immoral, "then there would appear to be a difficulty with the analysis of [the] essay" (201). He claims that his analysis would entail that contraception is wrong "only if something were denied a human future of value by contraception. Nothing at all is denied such a future by contraception, however" (201). Marquis considers and rejects four candidates

* Thanks are due to Jonathan Bennett, Frances Howard, Joel Kidder, and Steven Lee for helpful comments on an earlier draft of this paper.

[1] Don Marquis, "Why Abortion is Immoral," this JOURNAL, LXXXVI, 4 (April 1989): 183–202.

0022-362X/90/8705/268–277 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

for the role of subject of harm by contraception: (1) some sperm or other, (2) some ovum or other, (3) a sperm and an ovum separately, (4) a sperm and an ovum together. I shall concentrate on his treatment of (4), since I think it is clearly the most promising candidate.

Marquis offers the following argument in response to the suggestion that contraception deprives the combination of sperm and ovum of a valuable future-like-ours:

> At the time of contraception, there are hundreds of millions of sperm, one (released) ovum and millions of possible combinations of all of these. There is no actual combination at all. Is the subject of the loss to be a merely possible combination? Which one? This alternative does not yield an actual subject of harm either. Accordingly, the immorality of contraception is not entailed by the loss of a future-like-ours argument simply because there is no nonarbitrarily identifiable subject of the loss in the case of contraception (202).

I do not think it is clear from this passage what Marquis is claiming. It is all we get by way of argument against candidate (4), however, so we will have to work with it. I take it that Marquis does not mean that we could never tell which combination of sperm and ovum is the subject of loss in the case of successful contraception. I do not see how such an epistemic difficulty could rescue contraception from the charge of wrongdoing that Marquis's account levels against abortion. Consider the following two scenarios: (i) London is struck by a series of freak accidents involving power plants, which result in the deaths of almost everyone within a twenty-mile radius of Whitehall. The Prime Minister, Mrs. Butcher, is informed that there are a handful of survivors. She orders the army to round up the survivors, make a record of their names, and kill them. (ii) In response to the same information as in the previous example, Mrs. Butcher orders a hydrogen bomb to be detonated in London, thereby eliminating the possibility of discovering who survived the initial accidents. I hope it is clear that at least part of what is morally wrong with Mrs. Butcher's actions in each case is that some people are killed who would otherwise have lived. It makes no difference that in case (ii) it is impossible to tell which people were killed by the nuclear explosion.

It might be objected that contraception differs crucially from my example (ii) in that one cannot say of every use of contraception that it prevents conception. In most cases pregnancy would not have resulted anyway. One response to this would be to modify (ii) so that Mrs. Butcher is only told that there is a small but significant possibility that there are survivors. More importantly, though, I think it would be small comfort to the proponent of the moral acceptability of contraception to be told that contraception is morally permissible in all those cases in which it does not actually prevent conception.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The intuition that contraception is morally permissible is the intuition that it is permissible even, or perhaps especially, in those cases in which it fulfills its purpose.

Perhaps Marquis is not simply making an epistemic claim when he says that there is no nonarbitrarily identifiable subject of loss in the case of contraception. When he says that there are millions of possible combinations of sperm and ovum, but no actual combination, perhaps he means that there is no fact of the matter as to which sperm, if any, would have fertilized the ovum. It may be that the behavior of sperm is not strictly deterministic. Does this distinguish contraception from abortion and killing?

Consider the following scenario: two prisoners of conscience, Smith and Jones, are slated for execution in a small totalitarian republic. The president, Shrub, troubled by the effect of an Amnesty International campaign on his public image, decides to spare one of the prisoners. He cannot make up his mind whose life to spare, so he devises the following apparatus: Smith and Jones are placed in separate cells, each with air vents leading to a cannister of poison gas, which is set to release its contents at noon. A computer is programmed to select a three-digit number at random at one second before noon. If the number is even or zero, the computer will close the air vent in Smith's cell; if the number is odd, the computer will close the air vent in Jones's cell. The random-number selection process is truly indeterministic. The vice-president of the republic, Fowl, does not approve of Shrub bowing to liberal pressure. Fowl unplugs the computer at one minute before noon. The gas is released at noon, and both Smith and Jones die. I hope it is clear that Fowl has done something bad in this example. Two people have died instead of one. It is also clear that there is no fact of the matter as to which prisoner has been deprived of a valuable future by Fowl's actions. I do not think that the moral status of Fowl's action would change if the number of condemned prisoners was increased, but the number to be saved was kept at one. Even if Fowl's action makes it the case that a million and one die, instead of a million, he would have been responsible for the occurrence of one more death than would otherwise have occurred. In both of these cases, it seems that there is no nonarbitrarily identifiable subject of the loss of a valuable future. In both of these cases, Fowl's action is morally on a par with killing one person.

Let us return to the two-person case. We might still want to say that Fowl has harmed both Smith and Jones. We might say that the harm consists in lowering, from fifty percent to zero, their chances of surviving. (Indeed, given uncertainty about the length and quality of anyone's life, we should probably give a similar account of the harm involved in killing anyone.) If we increase the numbers, we can still

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

say that Fowl has harmed all the prisoners. In the case of a million and one prisoners, Fowl has reduced each of their chances of survival from just under one ten thousandth of one percent to zero. However we describe the immorality of Fowl's action, it is clear that what he does is wrong for precisely the same reason as a standard killing is wrong. Either Fowl harms all the prisoners, or the lack of a nonarbitrarily identifiable subject of loss does not provide a morally relevant distinction between this case and a standard case of killing.

The similarities between contraception and the example discussed above are clear in terms of the loss of a valuable future. Either contraception harms all the combinations of sperm and ovum, or the lack of a nonarbitrarily identifiable subject of loss does not provide a morally relevant distinction between contraception and abortion.

II

It might be claimed that there is a significant difference between the case of contraception and my poison-gas examples. In the poison-gas example, all the candidates for subject of the loss of a valuable future were things, people, in fact. In the case of contraception, the candidates I have been considering—combinations of sperm and ovum—are nonstarters, not because there is no nonarbitrarily identifiable subject of loss, but because a combination of sperm and ovum cannot be a subject of anything, because "it" is not a thing.[2] I do not know whether Marquis would be prepared to make this claim, but it is worth considering on its own merits.

It might be instructive at this point to consider what could be meant by 'a combination of sperm and ovum'. What has to be the case for such a thing to exist? It might be natural to assume that, for such a thing to exist, a sperm and an ovum would have to be combined. What would have to be the case for a sperm and ovum to be combined? Perhaps a sperm and an ovum would have had to have joined to form a zygote. Marquis might be taken to be espousing this reading of 'a combination of sperm and ovum' when he claims that there is no actual combination at the time of contraception. If this is his position, then he is, of course, right to say that there is no actual combination at the time of contraception (in the case of those forms of contraception which prevent a zygote from being formed, that is; it is an interesting question what Marquis would say about those forms of contraception which prevent the zygote from developing). If this is his position, though, why should we agree that his four categories of candidates for subject of harm are exhaustive? Why can we not say that the subject of harm is the mereological sum of a sperm and an ovum? Indeed, this is what I understand 'a combination of sperm and ovum' to mean. So what is the response to the

[2] This argument was suggested to me by Frances Howard.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

claim that a combination of sperm and ovum, understood as a mereological sum, is not a thing?

Perhaps the most obvious answer to the charge that a combination of sperm and ovum is not a thing is simply to deny it. I am inclined to pursue this option. To the extent that I am prepared to admit that a zygote or a fetus is a thing, I would claim that a combination of sperm and ovum is also a thing.

Let us assume, for the sake of argument, that a combination of a sperm and an ovum is not a thing. In that case, either the sperm and ovum, taken separately, *are* things, or neither is (I cannot imagine a plausible metaphysic that would classify one as a thing, but not the other). Consider the former alternative first. If both the sperm and the ovum are things, but the combination of the two is not, why can we not say that contraception harms both, because it deprives each one of the valuable future it would have had as a result of causal interaction with the other? This would appear to be Marquis's candidate (3), "a sperm and an ovum separately." Marquis rejects (3), because, "on this alternative, too many futures are lost. Contraception was supposed to be wrong, because it deprived us of one future of value, not two" (201). The obvious reply to this is that, on the assumption that a sperm and an ovum are both things, there *is* only one future lost as a result of contraception. It is the *same* future for both the sperm and the ovum. It is a shared future. Consider the possibility of human fusion. Imagine that it is possible for two people to fuse physically, so that a single person results, who has, among other things, some of the memories of both of the original people.[3] Imagine further that two people, Smith and Jones, will die, if they do not fuse with each other. Hector, an enemy of both Smith and Jones, knows of their predicament and of their plan to fuse. He locks them in adjoining rooms and waits until their anguished cries die out before gloating over their dead bodies. Given that Smith and Jones would have fused, if Hector had not locked the connecting door, it seems that Hector has deprived both Smith and Jones of a valuable future—the same valuable future. It might be objected that Hector has not deprived Smith and Jones of a valuable future, because they would have ceased to exist after they had fused, and a third person would have come into existence.[4] In this case, Hector has simply

[3] It might be objected that one cannot have a memory of, say, doing *x*, if one did not, oneself, do *x*. So, to say that the resulting person has some of the memories of both original persons might entail that the resulting person *is* both original persons. My argument does not require any such claim, so it is acceptable to modify the story so that it is merely claimed that the resulting person *seems* to have some of the memories of both original persons.

[4] It is important to note that, for Marquis, the central wrong-making feature of killing is that it deprives the victim of those valuable experiences which she would otherwise have had. A killing of someone who was about to die is not, therefore,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

prevented the existence of something with a valuable future. I do not think that such a distinction could possibly ground a judgment that Hector has not behaved at least as badly as if he had killed one person, who would otherwise have lived. However we describe Hector's action, it is clear that what he has done is morally on a par with killing. Could anyone honestly insist that it would be permissible for Hector to shoot Smith and Jones as they are running toward each other, about to fuse, but impermissible for him to shoot the person who results from the fusion a few seconds later?

III

The case of contraception, it might be argued, differs from my fusion example, because the sperm and the ovum are not things themselves. Before conception, goes the claim, there is just some stuff—fundamental particles or metaphysical simples or whatever—arranged in such a way that we are inclined to say that there is a sperm and an ovum. At conception, or thereabouts, all this stuff interacts in such a way that a thing comes into existence. This may seem like a drowning metaphysician clutching at straws, but I do not think that this, even if it is a good account of what there is, will ground a moral distinction between contraception and abortion.

Let us say, what is perhaps natural, that, in order to deprive something of a valuable future, that thing must at some time exist. So let us define an intransitive verb, 'to deprave' (not to be confused with the transitive verb 'to deprave'): to deprave is to act in such a way that some stuff, which would otherwise have interacted in such a way that a thing with a valuable future would have resulted, does not so interact. There would seem to be pairs of cases of contraception and abortion such that the contraception is a case of depraving, and the abortion is a case of depriving something of a valuable future. Some pairs of a depraving and a depriving differ only in that the hostile environment created by the contraceptive/abortifacient has the desired effect on the stuff before a thing comes into existence, in the case of the depraving, but on the thing after it has come into existence, in the case of the depriving. If there is a moral difference between a depraving and a depriving, it must be grounded in the fact that there is just some stuff that would otherwise have resulted in the existence of a thing, in the case of a depraving, but there actually *is* a thing, in the case of a depriving.

If we are to defend Marquis's position on the morality of abortion and contraception by appealing to the distinction between depriving and depraving, we must investigate further the category of "thing,"

morally bad, or at least does not share the central wrong-making feature of killings of people who would otherwise have lived long happy lives.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

which is at the heart of the distinction. It might be instructive to remind ourselves of Marquis's treatment of the anti-abortionist's attempt to base an argument on the category "human being":

> If 'human being' is taken to be a biological category, then the anti-abortionist is left with the problem of explaining why a merely biological category should make a moral difference. Why, it is asked, is it any more reasonable to base a moral conclusion on the number of chromosomes in one's cells than on the color of one's skin? If 'human being', on the other hand, is taken to be a moral category, then the claim that the fetus is a human being cannot be taken to be a premise in the anti-abortion argument, for it is precisely what needs to be established (186).

A similar argument can be used against an attempt to distinguish morally between contraception and abortion by appealing to the category "thing." Consider a case of depraving and a case of depriving something of a valuable future, which differ as little as is consistent with the former being definitely a case of depraving and the latter being definitely a case of depriving (the category of "thing" might have vague boundaries). In each case, there is some stuff that is affected by the contraceptive/abortifacient. In the latter case, but not in the former, there is also a thing that is the result of the interaction of the stuff. How might we establish that there is a thing in the latter case, but not in the former? If Marquis's position is to be defended, whatever makes the difference between thing and nothing must be definable without help from moral judgments on abortion and contraception, because those judgments are supposed to arise out of the difference between thing and no-thing. The most likely option would be to appeal to biological facts about the arrangement and interaction of the stuff in each case. But if our explanation of the difference between thing and no-thing depends simply on biological facts, it seems that we will still have to argue that the difference between a depraving and a depriving is morally significant.

IV

Marquis might object at this point that he does not have to argue that the difference between a depriving and a depraving is morally significant, because it is just obvious that there is a morally important difference between depriving a thing of its potentialities and preventing it from coming about that there is a thing with potentialities. It is difficult to know what to say to this, except to point out that it is far from obvious to me. Perhaps we could point out that a similar line of defense could be used by the pro-abortionists who argue that fetuses (at least in the early stages) cannot be victims. It is just obvious, we might claim, that an early fetus cannot be a victim, because it lacks sentience. Alternatively, we might claim that it is just obvious that there is a morally significant difference between depriving a sentient victim of a valuable future and depriving a

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

presentient victim of a valuable future, on the other. Of course, we might also claim that it is obvious that there is a morally significant difference between depriving something that has not been born yet of a valuable future and depriving something that has been born of a similar future. We might disagree with Marquis's claim that the central wrong-making feature of a killing is that it deprives something of a valuable future. On the one hand, we might claim instead that the central wrong-making feature of a killing is that it deprives a certain specified sort of something of a valuable future. On the other hand, we might claim that the central wrong-making feature of a killing is what a depriving and a depraving have in common. I have tried to show that the distinction between thing and no-thing is no more morally significant than the distinction between sentient thing and presentient thing. Of course, one can always claim that a particular intuition is just rock-bottom. But if that is Marquis's claim about the distinction between thing and no-thing, then I would claim that he has given no arguments against abortion that would appeal to any of those who do not share his intuitions. My argument is not, however, that we should simply pick whichever distinction we feel most strongly about, and then base our position around that; at least not with respect to the question of what is bad about depriving something of a valuable future. I would claim that a consideration of the various distinctions—between postnatal and prenatal thing, between sentient and presentient thing, between thing and no-thing—should lead us to reject the claim that any of them makes a moral difference with respect to what is bad about depriving something of a valuable future (and, of course, what is bad about preventing it from coming about that there is a thing with a valuable future).

V

Another approach to the question of whether we should imbue the distinction between thing and no-thing with moral significance would be to ask why it is bad to deprive something of a valuable future. As far as I can see, there are two broad approaches that can be taken to answer this question (other than simply insisting that it *is* bad). On the one hand, we could claim that it is bad to deprive something of a valuable future, because it is bad, other things being equal, to prevent the occurrence of the valuable experiences and the like that constitute a valuable future. I hope it is clear why this approach will militate against drawing a morally significant line between depraving and depriving. On the other hand, we could claim that it is bad to deprive something of a valuable future, because the victim is, in some sense, entitled to her valuable future. We might say that she has a right to her valuable future. This would seem to allow us to draw a line between depraving and depriving. Depriving something of a valuable future violates certain rights of that thing. A

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

depraving, on the other hand, involves no violations of rights, because there is no victim. The right to life, though, is far from unproblematic. Why does anything have a right to a valuable future, as opposed to, say, a very bad future, or no future at all? Why does this fetus have a right to the future that would ensue if I do not perform the abortion, rather than the future that would ensue if I do?[5] We cannot answer that question with an appeal to the supposed immorality of performing an abortion, since it is that very immorality that we are seeking to establish by appeal to the right to a valuable future. How, in general, can we explain why something has a right to the good aspects of its future, but not to the bad ones? Perhaps we could appeal to the goodness of such good aspects obtaining (goodness, that is, that is not dependent on the claim that the possessor of such a future has a right to those good aspects). If my right to a valuable future is grounded in the value of the occurrence of those experiences and the like that constitute such a future, however, the very same grounding will apply to the immorality of a depraving.

Perhaps it could be argued that my right to a certain sort of future is not restricted to a valuable future. In fact, I have a right to all the unpleasantness that might be in my future, too. It is a common feature of rights that they can be waived, and so I am taken to have waived my rights to the unpleasant aspects of my future. Thus, it is not true to say of a doctor, who has deprived me of much pain and suffering, that she has violated my rights. If we take this line, it is very hard to see what is being claimed as the content of something's right to a future. If my right is not restricted to valuable aspects of my future, just what do I have a right to? I presume the claim is not that I have a right to my *actual* future, because, if that were the case, no actual behavior that affected me could violate such a right. Perhaps the content of my right varies, according to the situation I am in. Perhaps, in general, I have a right to whatever future would ensue, if the agent who might act so as to affect me does not so act. But this clearly will not do. If an agent's action will affect me, then his not performing the action will also affect me, since my future will be different from how it would otherwise have been.

Perhaps we will have to postulate a distinction between *doing* and *allowing*, and claim that I have a right to the future that would have ensued, if the agent had not done the thing in question. Quite apart

[5] It might sound strange to talk of the future that would ensue for the fetus if I perform an abortion. The fetus, it would seem, will have no future if I perform an abortion. I think it is clear, however, that the option of nonexistence for a fetus, or an adult human being, can be compared with other possible futures. It is also clear that nonexistence is preferable to some futures, futures filled with unmitigated pain and misery, for example. If we ask what sort of a future a fetus (or any other being) has a right to, it would be strange to exclude nonexistence from the options being considered.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

from the difficulty of drawing a distinction between doing and allowing (or anything remotely resembling such a distinction), and the much greater difficulty, if not impossibility, of establishing that such a distinction can bear any moral weight, this distinction does not accord with our normal intuitions about rights. Any version of the doing/allowing distinction will have to place the neglect of a new-born baby, so that it starves to death, on the side of allowing. But it would be a strange position indeed that condemned abortion as the violation of the fetus's right to a valuable future, but did not render the same judgment about allowing a new-born to starve to death. Perhaps Marquis could claim that a thing that might have a valuable future just does have a right to a valuable future, and that is all there is to it. I would claim that, if we are disinclined to accept such a right as a basic intuition and seek some grounding for the right, we will have a hard time finding grounding for such a right which will not also ground the immorality of contraception.

VI

I have attempted to give reasons why we should be, at the very least, uncomfortable with allowing moral weight to the distinction between thing and no-thing, and thus the distinction between abortion and contraception. Lest I be taken to be arguing against contraception (and, indeed, sexual abstinence), I should add that I do not consider the deprivation of a valuable future to be the only, or even the central, wrong-making feature of a standard killing. Thus, although what is bad about depriving someone of a valuable future is shared by abortion, regular use of contraception, and sexual abstinence, there are many morally significant factors that distinguish standard killings from abortions, and abortions from contraception. But they are the subject of another paper.[6]

I have argued in this paper that Marquis fails to distinguish morally between contraception and abortion. I have also argued that an attempt to distinguish between contraception and abortion by appeal to a parsimonious ontology does not provide a morally relevant distinction. I conclude that Marquis is unable to distinguish morally between contraception and abortion without appealing to morally relevant features other than what he calls the "wrong-making feature of one's being killed." I conjecture that any appeal to morally relevant features sufficient to ground a moral distinction between abortion and contraception will also ground a moral distinction between abortion and standard cases of killing.

ALASTAIR NORCROSS

Syracuse University

[6] Although I do not wish to wade into this topic here, I shall say that I consider the main weakness of Marquis's account of the wrongness of a standard killing to be that it does not take into consideration the full range of consequences of such an act.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

 Digitized by Arya Samaj Foundation Chennai and eGangotri

NOTES AND NEWS

Behaviorism is changing its name, after 17 years, to *Behavior and Philosophy* in order to reflect more adequately changes that have taken place in recent years in the editorial scope of the journal. The journal welcomes papers in any area of philosophy which illuminate either directly or indirectly the theory, practice, or technology of behavioral and psychological science. Papers in philosophy of mind are especially welcome. Submissions, in triplicate, may be addressed to George Graham, Editor, *Behavior and Philosophy,* Phil. Dept., Univ. of Alabama, Birmingham, AL 35294.

The International Symposium on "Structures in Mathematical Theories," organized by the Department of Logic and Philosophy of Science at the University of the Basque Country, will take place in San Sebastian, Spain, September 25–29. The aim of the Symposium is to provide for discussion among mathematicians, philosophers of science, and historians of science to contribute to a better understanding of five major areas: Mathematical Theories and Empirical Theories, Applications of Mathematical Theories, History and Sociology of Mathematical Theories, Methods of Research into Mathematics, and Structures of Mathematical Theories. Papers may be submitted to Javier Echeverria, Depto. de Logica y Filosofia de la Ciencia, Apdo. 1249, E-20080 San Sebastian, Spain. For further information, contact Andoni Ibarra, Depto. de Logica y Filosofia de la Cienca, Facultad de Filosofia y CCEE, Apdo. 1249, E-20080 Donostia-San Sebastian, Spain.

The Society for the Study of Ethics & Animals will met in conjunction with the Division meetings of the American Philosophical Association. Papers are hereby invited (by May 20) for the 1990 Eastern meeting in Boston in late December, and for all 1991 meetings. Papers are welcome on any topic connected with ethical issues affecting nonhuman animals. To submit papers or for more information, you may write to Harlan B. Miller, Phil. Dept., Virginia Polytechnic Institute and State University, Blacksburg, VA 24061.

The Department of Philosophy at the University of California/Berkeley has established a series of annual lectures to honor the late George Myro, who died in December of 1988. The first four lecturers, in March 1989, were George Bealer, Sally Haslanger, Stephen Yablo, and Mark Bedau, all of whom had been students of Prof. Myro. Friends and students of Prof. Myro who wish to contribute to the endowment of the series may do so by sending donations payable to the George Myro Memorial Fund, Phil. Dept., Moses Hall, Univ. of California, Berkeley, CA 94720.

The Department of Philosophy of the University of Illinois/Chicago is pleased to announce the appointment of Charles W. Mills as Assistant Professor, effective September.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVII, NO. 6, JUNE 1990

THE STRUCTURE AND CONTENT OF TRUTH*

NOTHING in the world, no object or event, would be true or false if there were not thinking creatures. John Dewey, in whose honor and memory the lectures that constituted this essay were given, drew two conclusions: that access to truth could not be a special prerogative of philosophy, and that truth must have essential connections with human interests. He was contemptuous of the philosophical tradition that viewed truth as correspondence between thought and a reality inaccessible to experimental research and ordinary practice. He believed this picture of truth was designed to serve the thesis that philosophers possess a privileged technique for achieving a form of knowledge different from, and superior to, science. Dewey[1] wrote that

> . . . the profuseness of attestations to supreme devotion to truth on the part of philosophy is matter to arouse suspicion. For it has usually been a preliminary to the claim of being a peculiar organ of access to highest and ultimate truth. Such it is not. . . . Truth is a collection of truths; and these constituent truths are in the keeping of the best available methods of inquiry and testing as to matters-of-fact; methods, which are, when collected under a single name, science. As to truth, then, philosophy has no pre-eminent status . . . (*ibid.*, p. 410).

Dewey's aim was to bring truth, and with it the pretensions of philosophers, down to earth. We may with justice feel that Dewey confused

* Presented as three lectures on "The Concept of Truth," given at Columbia University in November 1989; the first, "The Structure of Truth," on November 9; the second, "Truth and Knowledge," on November 16; and the third, "The Contents of Truth," on November 20. These lectures, made possible by the John Dewey Foundation, constitute the sixth series of John Dewey Lectures, which were established in 1967 to honor the late John Dewey, who was a professor of philosophy at Columbia from 1905 to 1930. I am grateful to Akeel Bilgrami, Ernest LePore, Isaac Levi, and W. V. Quine for helpful suggestions and friendly encouragement.

[1] *Experience and Nature* (New York: Dover, 1958).

0022-362X/90/8706/279–328 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the question what sort of concept truth is with the question what kinds of truth there are. But it is clear that the two issues are related, since what falls under the concept obviously depends on what the concept is. And the idea of ensuring that the domain of truth can be convincingly brought within the scope of human powers by cutting the concept down to size is hardly unique to Dewey; Dewey saw himself as sharing the views of C. S. Peirce and William James in this matter, and in one way or another the basic theme reappears today in the writings of Hilary Putnam, Michael Dummett, Richard Rorty, and many others.

Those who wish to debunk or deflate the concept of truth often start by rejecting any hint of the correspondence theory, but Dewey[2] saw no harm in the idea of correspondence as long as it was properly understood. "Truth means, as a matter of course, agreement, correspondence, of idea and fact," he said, but immediately went on, "but what do agreement, correspondence mean?" (*ibid.*, p. 304). He answered, "the idea is true which works in leading us to what it purports" (*ibid.*), and he quotes James[3] with approval:

> . . . any idea that will carry us prosperously from any one part of experience to any other part, linking things satisfactorily, working securely, simplifying, saving labor, is true for just so much, true in so far forth (*ibid.*, p. 58).

Probably few philosophers are now tempted by these swinging, sweeping formulations. But the problem the pragmatists were addressing—the problem of how to relate truth to human desires, beliefs, intentions, and the use of language—seems to me the right one to concentrate on in thinking about truth. It also seems to me this problem is not much nearer a solution today than it was in Dewey's day.

To see this as the main problem about truth—or indeed as a problem at all—is to assume that the concept of truth *is* related in important ways to human attitudes; something it is not uncommon to

[2] *Essays in Experimental Logic* (New York: Dover, 1953)

[3] *Pragmatism* (New York: Longmans & Green, 1907). Elsewhere [*Logic: The Theory of Inquiry* (New York: Holt, 1938)], Dewey says:

> The best definition of *truth* from the logical standpoint which is known to me is that of Peirce: "The opinion which is fated to be ultimately agreed to by all who investigate is what we mean by the truth" (p. 58).

But usually Dewey was closer to James: ideas, theories, are true if they are "instrumental to an active reorganization of the given environment, to a removal of some specific trouble and perplexity. . . . The hypothesis that works is the *true* one"—*Reconstruction in Philosophy* (New York: Holt, 1920), p. 156.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

doubt. It is not uncommon, in fact, to doubt whether the concept of truth is of any serious philosophic importance at all.

Rorty captures Dewey's intention of removing truth from a realm so exalted only philosophers could hope to attain to it when he introduces his *Consequences of Pragmatism*[4] with the words:

> The essays in this book are attempts to draw consequences from a pragmatist theory about truth. This theory says that truth is not the sort of thing one should expect to have a philosophically interesting theory about . . . there is no interesting work to be done in this area (*ibid.*, pp. xiii–xiv).

But it seems to me Rorty misses half the point of Dewey's attitude toward the concept of truth: Dewey says that *truths* are not in general the special province of philosophy; but he also insists that *truth* is what works. This is not the same as the thesis that there is nothing interesting to be said about truth. Dewey found plenty that was interesting to say about what works.

Rorty[5] has compared my views on the nature of truth with Dewey's. I find much of what he has to say on this topic congenial and penetrating, and I think he is right that in a general way I share Dewey's attitude toward truth. In one respect, though, a respect on which I just touched, Rorty may have us both wrong; as I read him, Dewey thought that once truth was brought down to earth there were philosophically important and instructive things to say about its connections with human attitudes, connections partly constitutive of the concept of truth. This is also my view, though I do not think Dewey had the connections right.

Rorty correctly notes the fundamental role I assign to Alfred Tarski's work as providing a way of discussing the understanding of language, and he sees clearly that for me this is related to the rejection of a representational picture of language and the idea that truth consists in the accurate mirroring of facts. These are matters to which I shall turn presently. In this essay, I first discuss the notion, often associated with Tarski's approach, that talk of truth is essentially redundant, and has no important properties beyond those specified in Tarski's definitions of truth. The first section ends with a defense of the claim that Tarski's definitions may legitimately be treated as conveying substantive truths about a language, but that in

[4] Minneapolis: Minnesota UP, 1982.
[5] "Pragmatism, Davidson and Truth," in LePore, ed., *Truth and Interpretation* (New York: Blackwell, 1986), pp. 333–355. Also see his "Representation, Social Practise, and Truth," *Philosophical Studies*, xxx (1988): 215–228.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

this case there must be more to the concept than Tarski specified. In the second section of the essay, I turn to various attempts to say what more is involved: I discuss correspondence theories, coherence theories, and theories that in one way or another make truth an epistemic concept. I reject all these kinds of theories. In the third section, I propose an approach that differs from the rest, one that makes the concept of truth an essential part of the scheme we all necessarily employ for understanding, criticizing, explaining, and predicting thought and action.

I

The redundancy theory best fits phrases like 'It is true that' or 'It is a fact that' when prefixed to a sentence. Such phrases may be regarded as truth-functional sentential connectives which, when added to a true sentence, yield a true sentence, and when added to a false sentence, yield a false sentence. These connectives would then function exactly like double negation (when negation is classically conceived). At least so far as cognitive content and truth conditions are concerned, such appendages are redundant.

Frank Ramsey[6] seems to have thought all uses of the concept of truth are like this. He says: " 'It is true that Caesar was murdered' means no more than that Caesar was murdered" (*ibid.*, p. 143). He then considers cases like 'Everything he says is true' from which reference to truth is not so easily eliminated, and suggests that, if we restrict ourselves to propositions of the form *aRb*, we could render 'Everything he says is true' as 'For all *a*, *R*, *b*, if he says *aRb*, then *aRb*'. Ramsey adds that, if all forms of proposition are included, things get more complicated, "but not essentially different" (*ibid.*). Although Ramsey does not always clearly distinguish between propositions and sentences, or the use of sentences and their mention, one gets the impression that, if Ramsey had carried out the "more complicated" analysis, he might have ended up with something much like one of Tarski's truth definitions. In any case, Ramsey thought he had said enough to show that "there is really no separate problem of truth but merely a linguistic muddle" (*ibid.*, p. 142).[7]

Ramsey was wrong if he thought the truth-functional-connective analysis of the use of 'true' could be directly applied to sentences like 'Everything he says is true', for in the former case the truth phrase is

[6] "Facts and Propositions" (1927), reprinted in *The Foundations of Mathematics* (New York: Humanities, 1931), pp. 138–155.

[7] P. F. Strawson says much the same in his famous debate with J. L. Austin, in "Truth," *Proceedings of the Aristotelian Society, Supp. Vol. XXIV* (1950): 129–156.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

viewed as a connective, whereas in the latter case it must be treated as a predicate and, if we follow Tarski, it must belong to a different language from the language of the sentences of which it is predicated. It might be possible to treat phrases like 'It is true that' as predicates of propositions rather than as sentential connectives, but again the redundancy would be far less manifest than Ramsey claimed.

Many philosophers have nevertheless regarded Tarski's work as essentially a matter of straightening out Ramsey's insight. W. V. Quine,[8] for example, writes: "To say that the statement 'Brutus killed Caesar' is true . . . is in effect simply to say that Brutus killed Caesar," and he tells us, in a footnote, to see Tarski for the "classic development" of this theme (*ibid.*, p. 24). Putnam maintains that Rorty and Quine share this view of truth. According to Putnam,[9] Rorty and Quine believe that "[t]o call a sentence 'true' is not to ascribe a property, truth, to a sentence; it is just another way of asserting the sentence" (*ibid.*, p. 62). (He adds that this is called the "disquotational view"—"in the jargon of Davidsonian philosophers of language" (*ibid.*). Maybe so, but then I am not a Davidsonian, for I am not tempted to refer to Tarski's truth definitions as "disquotational.") In any case, Putnam is not endorsing this thesis; he is attacking it as "purely formal" and "empty."

It is not clear to me whether Putnam thinks Tarski's work on truth is no more than a technical improvement on what is basically a redundancy theory, but others have certainly taken this line. Stephen Leeds[10] has suggested that the "usefulness" or importance of the concept of truth might just consist in this, that it gives us a way to say things like "Most of our beliefs are true," where we want to talk of, or perhaps assert, an infinite or otherwise unlistable set of sentences. Ramsey did not explain how to do this; Tarski did. Paul Horwich,[11] like Leeds, considers Tarski to be a redundancy theorist; Horwich is persuaded that, despite our intuition that truth is a central and important concept, "the notion of truth was completely captured by Tarski" (*ibid.*, p. 192). This idea, that Tarski did all that could be done for the concept of truth, Horwich calls the deflationary theory of truth.

Although he does not agree with Horwich that truth as Tarski defined it specifies truth conditions adequate to an account of what

[8] *Word and Object* (Cambridge: MIT, 1960).
[9] "A Comparison of Something with Something Else," *New Literary History*, XVII (1985): 61–79.
[10] "Theories of Reference and Truth," *Erkenntnis*, XIII (1978): 111–130.
[11] "Three Forms of Realism," *Synthese*, LI (1982): 181–201.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

language users know, Scott Soames[12] concurs in calling Tarski's approach to truth deflationary, and like Horwich, he thinks that, when it comes to explicating the concept of truth, we should not ask for anything more, aside from the application of truth to propositions, etc.

Hartry Field,[13] in a useful article, explores the case for a deflationary concept of truth, and shows how hard it would be to go beyond it. He explains what he means by calling a theory of truth deflationary in approximately the way Horwich does: truth is disquotational and nothing more; but he is less certain than Horwich that Tarski should (or must) be seen as a disquotationist, though he believes Tarski's work can be appropriated by the disquotationist. Michael Williams[14] has recently characterized the views of deflationists this way: they

> . . . think that when we have pointed to certain formal features of the truth-predicate (notably its 'disquotational' feature) and explained why it is useful to have a predicate like this (e.g. as a device for asserting infinite conjunctions), we have said just about everything there is to be said about truth (*ibid.*, p. 424).

He explicitly accepts a deflationary attitude to truth.[15]

How plausible are these various deflationary theories of truth? If we restrict the redundancy theory to occurrences of 'true' as part of a truth-functional sentential connective (as in 'It is true that snow is white'), then it is clear that such uses play only a small role in our talk of truth; this cannot be the whole story. Can disquotational theories do better? Tarski's truth definitions are disquotational in this sense: given the definition (and set theory and formal syntax), and given a sentence in the form " 'Snow is white' is true," we can prove that the sentence 'Snow is white' is equivalent. Thus, the sentence in which 'Snow is white' was only mentioned is provably equivalent to the sentence 'Snow is white' itself; the original " 'Snow is white' " has been stripped of its quotation marks; removing the quotation marks cancelled out, so to speak, the truth predicate. And even when we cannot remove the quotation marks because there are no quotation marks to remove (as in 'Everything he said was true' or 'A valid rule of inference guarantees that from true premises only true conclu-

[12] "What is a Theory of Truth?" this JOURNAL, LXXXI, 8 (August 1984): 411–429.

[13] "The Deflationary Conception of Truth," in C. Wright and G. McDonald, eds., *Facts, Science and Morality* (New York: Blackwell, 1987), pp. 55–117.

[14] "Epistemological Realism and the Basis of Skepticism," *Mind*, XCVII (1988): 415–439.

[15] See "Do We (Epistemologists) Need a Theory of Truth?" *Philosophical Topics*, XIV (1986): 223–242.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

sions follow'), Tarski has shown us how to get rid of the truth predicate, since it has been explicitly defined.[16] This makes clear that Tarski's truth definitions are not strictly disquotational, since they do not depend on stripping the quotation marks from individual sentences in order to eliminate truth predicates. Still less do they depend on using the actual sentences said to be true to effect the elimination; this is obvious when the definition of truth for one language is given in another. One cannot find an English equivalent of the English sentence " 'Schnee ist weiss' is true (in German)" simply by removing the quotation marks from " 'Schnee ist weiss'."

Still, the fact remains that Tarski's methods allow us to replace the truth predicates he defines in any context, and the replacement leaves no explicitly semantical predicates in its wake; in this respect, his truth predicates are like the sentential connective 'it is true that', which may be removed by simple deletion. What is striking, of course, is not that the phrase 'is true' can be replaced, for that can be the point of definition; what is striking is that it is not replaced by anything else, semantic or otherwise. It is presumably this feature which leads Putnam to say that, according to such theories, truth is not a property. (This cannot be exactly right as applied to Tarski's truth definitions, however. Tarski's truth predicates are legitimate predicates, with an extension no predicate in the object language has. But one sees the point of Putnam's remark.) Putnam concludes that Tarski's truth predicates have nothing to do with semantics or the common conception of truth: "As a philosophical account of truth, Tarski's theory fails as badly as it is possible for an account to fail" (*op. cit.*, p. 64).

What is clear is that Tarski did not define the concept of truth, even as applied to sentences. Tarski showed how to define a truth predicate for each of a number of well-behaved languages, but his definitions do not, of course, tell us what these predicates have in common. Put a little differently: he defined various predicates of the form '*s* is true$_L$', each applicable to a single language, but he failed to define a predicate of the form '*s* is true in *L*' for variable '*L*'. The point was made by Max Black[17] and subsequently by Dummett;[18] but of course Tarski had made this thunderously clear from the start by

[16] This point, often credited to Leeds, was made by Tarski in "The Semantic Conception of Truth," *Philosophy and Phenomenological Research*, IV (1944), p. 359. Tarski also notes that mere disquotation cannot eliminate the word 'true' from sentences like 'The first sentence written by Plato is true'. (But neither has Tarski shown how to eliminate this use of the truth predicate unless he has a definition of truth for the language Plato spoke.)

[17] *Language and Philosophy* (Ithaca: Cornell, 1949), p. 104.

[18] "Truth," in *Proceedings of the Aristotelian Society*, LIX (1958–9): 141–162.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

proving that no such single predicate could be defined in a consistent language, given his assumptions concerning truth predicates. Given those constraints, there was never any chance that he could give a general definition of the concept of truth, even for sentences. If we consider the application of the concept of truth to beliefs and related phenomena like claims and assertions, it is obvious in a further way that Tarski did not attempt a really general definition. Considering how evident it is that Tarski did not give a general definition of truth, and the fact that perhaps his most important result was that it could not be done along lines that would satisfy him, it is remarkable how much effort some critics have put into the attempt to persuade us that Tarski failed to provide such a definition.

Dummett says in the "Preface" to *Truth and Other Enigmas*[19] that the "fundamental contention" of his early article "Truth" was that any form of the redundancy theory (and he includes Tarski's truth definitions in this category) must be false because no such theory can capture the point of introducing a truth predicate. This can be seen, he argues, from the fact that, if we have a Tarskian truth definition for a language that we do not understand,

> we shall have no idea of the point of introducing the predicate . . . unless . . . we already know in advance what the point of the predicate so defined is supposed to be. But, if we do know in advance the point of introducing the predicate "true" then we know something about the concept of truth expressed by that predicate which is not embodied in that . . . truth definition (*ibid.*, pp. xx–xxi).

Dummett adds that "[a]lthough this contention was so obvious when formulated I believe that it was worth stating at the time" (*ibid.*). He is right: the contention was obvious, and was worth stating, at least to me.[20] The application to theories of meaning is important; but the issue is more general: Tarski knew he could not give a general definition of truth, and so there was no formal way he could capture "the point" of introducing truth predicates, whether that point concerned the connection between truth and meaning or between truth and some other concept or concepts.

[19] London: Duckworth, 1978.

[20] My confusion on this point is most apparent in "Truth and Meaning," in *Inquiries into Truth and Interpretation* (New York: Oxford, 1984). My mistake was to think we could *both* take a Tarski truth definition as telling us all we need to know about truth *and* use the definition to describe an actual language. But even in the same essay I (inconsistently) discussed how to tell that such a definition applied to a language. I soon recognized the error. (See the "Introduction," pp. xiv–xv, and other essays in *Inquiries into Truth and Interpretation*.)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Dummett and others have attempted in various ways to make the slow-witted among us appreciate the failure of Tarski's truth predicates to capture completely the concept of truth. The central difficulty, as we have seen, is due simply to the fact that Tarski's definitions give us no idea how to apply the concept to a new case, whether the new case is a new language or a word newly added to a language [these are really the same point, put both ways by Dummett (*op. cit.*) and the second way by Hartry Field[21]]. This feature of Tarski's definitions can in turn easily be traced to the fact that they depend on giving the extension or reference of the basic predicates or names by enumerating cases: a definition given in this way can provide no clue for the next or general case.

A number of criticisms of, or comments on, Tarski's treatment of truth depend on the enumerative aspect of his definitions. One such is the claim that Tarski's definitions cannot explain why, if the word 'snow' had meant "coal," the sentence 'Snow is white' would have been true if and only if snow had been black. Putnam and Soames both make the point, but for Putnam it is a criticism, while for Soames it illustrates the folly of expecting much from a theory or definition of truth. Another complaint is that Tarski's definitions do not establish the connection between truth and meaning that many philosophers hold to be essential. (Again, for Putnam this shows there is something basically wrong with Tarski's conception of truth; for Soames it is one more example of the laudably deflationary aspect of Tarski's definitions.) A closely related comment is that Tarski does not relate truth to the use or users of language (Field, Putnam, Soames, Dummett). Whatever the value of these remarks may be, it is worth keeping in mind that they all trace back to the same simple feature of Tarski's work: by employing a finite and exhaustive list of basic cases in the course of defining satisfaction (in terms of which truth is defined), he necessarily failed to specify how to go on to further cases.

Despite the limitations that have been identified or imagined in Tarski's work on truth, a number of philosophers, as we have seen, have endorsed that work as embracing all of truth's essential features. These philosophers include Rorty, Leeds, Michael Williams, Horwich, Soames, and, according to Putnam, Quine; also, according to Rorty, me.[22]

[21] "Tarski's Theory of Truth," this JOURNAL, LXIX, 13 (July 13, 1972): 347–375.

[22] Not surprisingly, the views of the people on this list differ on the sense in which Tarski is a deflationist. Horwich, for example, introduced the term 'deflationist' in speaking of Tarski, but he holds that Tarski's "schema" gives the truth conditions, and hence the meanings, of the expressions of a language; his view is essentially that

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

I do not belong on this list, however. The basic argument, which was intended to reveal Tarski as a deflationist, can be taken two ways: as showing that he did not capture essential aspects of the concept of truth, or as showing that the concept of truth is not as deep and interesting as many have thought.[23] Like Dummett and Putnam, I think we must take it in the first of these two ways. The reason is plain. Nothing in Tarski's truth definitions hints at what it is that these definitions have in common. Unless we are prepared to say there is no single concept of truth (even as applied to sentences), but only a number of different concepts for which we use the same word, we have to conclude that there is more to the concept of truth—something absolutely basic, in fact—which Tarski's definitions do not touch. What is mildly puzzling is that some philosophers who appeal to a version of the basic argument to show that Tarski's truth predicates are deflationary at the same time accept a deflationary theory. But if the basic argument is sound, it shows that definitions like Tarski's, or theories built on the same lines, cannot capture the concept of truth.

There is a further claim or assumption about Tarski's work which, though it is often run together with some of the points just rehearsed, deserves separate discussion. The theme is that, if we accept one of Tarski's truth definitions, then statements that ought, if truth were properly characterized, to be empirical statements are turned into truths of logic. Thus, according to Putnam, a sentence like " 'Schnee ist weiss' is true (in German) if and only if snow is white" ought to be a substantive truth about German, but if for the predicate '*s* is true (in German)' we substitute a predicate defined in Tarski's style, the apparent substantive truth becomes a truth of logic.[24] It is easy to see that whatever there is to this argument depends on the same feature of Tarski's method we have been discussing: if the extension of a predicate is defined by listing the things to which it applies, applying the predicate to an item on the list will yield a statement equivalent to a logical truth. (For technical reasons

---

of my "Truth and Meaning." Most of the others think that Tarski's deflationist approach shows that truth as he defines it has nothing to do with meaning.

[23] The first attitude stands out in Putnam's remark that the property Tarski defined is not "even doubtfully or dubiously 'close' to the property of truth. It just isn't truth at all"—*op. cit*, p. 64. Soames represents the second view: "What does seem right about Tarski's approach is its deflationary character." But "Tarski's notion of truth has nothing to do with semantic interpretation or understanding" —"What is a Theory of Truth?" pp. 429, 424.

[24] For versions of this argument, see Putnam, *op. cit.*, and "On Truth," in Leigh Cauman et alia, eds., *How Many Questions?* (Indianapolis: Hackett, 1983), pp. 35–56.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

this is an oversimplified explanation of this aspect of Tarski's method when the object language includes quantifiers, etc. The force of the point remains.[25]) This seems to be Putnam's main reason for saying that Tarski failed "as badly as it is possible to fail" in giving a philosophical account of truth. Soames may be thinking along the same lines when he maintains that the only way to defend Tarski's philosophical interpretation of his work is to reject the demand that applications of his truth and satisfaction predicates have empirical content. To meet the demand would, Soames says, be "incompatible" with Tarski's work (*op. cit.*, p. 425).

The argument is spelled out at some length by John Etchemendy (*op. cit.*). According to Etchemendy, Tarski's goal was to formulate predicates with two properties: first, they should be related in a specific way to the intuitive concept of truth, and, second, they should be guaranteed, as far as possible, against the threat of paradox and inconsistency. The first condition was met by devising a concept that could easily be shown to apply to all and only the true sentences of a language. The relation to the intuitive concept of truth is made manifest by convention-T. Convention-T requires that the truth predicate '*s* is true$_L$' for a language *L* be so characterized as to entail, for every sentence *s* of *L*, a theorem of the form '*s* is true$_L$ if and only if *p*', when '*s*' is replaced by a systematic description of *s* and *p* is replaced by a translation of *s* into the language of the theory. Let us call these theorems T-sentences. The predicate in T-sentences, '*s* is true$_L$', is a one-place predicate; the subscript is not a variable, but the name or description of a particular language and an undetachable part of the predicate. The relation to the ordinary concept of truth is apparent from the fact that T-sentences remain true if for the Tarski-style truth predicate we substitute the English predicate '*s* is true in *L*'. (This is a two-place predicate: we can substitute names or descriptions of other languages for '*L*'.) The demand that the truth predicate not threaten to introduce inconsistencies into the theory or language is met by giving an explicit definition of the predicate using no semantic concepts; thus, any challenge to consistency such concepts might present has been avoided. If the metalanguage is consistent before the introduction of the truth predicate, it is guaranteed to remain so after the introduction.

T-sentences containing Tarski's truth predicates *seem* to convey substantive facts about the object language, namely, that its sentences are true under conditions specified by the T-sentence

[25] For the development of this theme, see the works of Putnam referred to in the last footnote; also Soames, *op. cit.*; and John Etchemendy, "Tarski on Truth and Logical Consequence," *The Journal of Symbolic Logic*, LIII (1988): 51–79.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

('Schnee ist weiss' is true in German if and only if snow is white), but in fact, says Etchemendy, "they carry no information about the semantic properties of [the] language, *not even about the truth-conditions of its sentences*" (his emphasis, *op. cit.*, p. 57). The reason for this is that T-sentences are truths of logic, and so cannot tell us anything logic alone cannot tell us. T-sentences are truths of logic, in turn, because they follow from Tarski's definitions, and these are merely stipulations; we are misled because of "the ease with which we read substantive content into what is intended as a stipulative definition, the ease with which we replace the 'if and only if' of definition with the 'if and only if' of axioms or theorems" (*op. cit.*, p. 58). If we want to state substantive facts about a language, we must substitute in T-sentences and elsewhere a predicate that conveys something like the intuitive concept of truth. If we do this, "the claims we make will sometimes look strikingly like clauses" in Tarski's definitions and (if correct) yield genuine information about the semantic properties of a language.

But, and this is Etchemendy's central message, the two enterprises—of defining truth according to Tarski's aims, and of providing a formal but substantive semantic account of a language—are not only totally different enterprises, but are in "quite direct opposition to one another. . . . For without setting aside Tarski's principal goal, there is a sense in which semantics simply cannot be done" (*op. cit.*, pp. 52–3). The difference between the two is that the first demands a predicate that can be eliminated without remainder from all contexts, while the second requires a "fixed, metatheoretical" notion of truth. Employing the second concept would directly defeat the point of Tarski's project. Thus, the relation between Tarski's intended and successful achievement, on the one hand, and the project of supplying a way of describing the semantics of interpreted languages, on the other, is "little more than a fortuitous accident" (*op. cit.*, pp. 52–3).

Putnam, Soames, and Etchemendy agree that Tarski's T-sentences only appear to state empirical truths about a language; they are in fact "tautologies" (Putnam). They differ in their appraisals of the thesis on which they agree: Putnam[26] thinks what Tarski defined "just isn't truth at all"; Soames and Etchemendy claim Tarski did what he set out to do. Soames holds that Tarski was right to give a deflationary account of truth, while Etchemendy thinks empirical semantics is a legitimate study which Tarski was not pursuing.

[26] "A Comparison of Something with Something Else," p. 64.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

How should we think about these claims? One thing is certain: Tarski did not agree with these assessments of his results. In "The Semantic Conception of Truth,"[27] there is a section headed "Conformity of the Semantic Conception of Truth with Philosophical and Common-Sense Usage." Let me quote from it:

> As far as my own opinion is concerned, I do not have any doubts that our formulation does conform to the intuitive content of that of Aristotle . . . some doubts have been expressed whether the semantic conception does reflect the notion of truth in its common-sense and every-day usage. I clearly realize . . . that the common meaning of the word "*true*"—as that of any other word of everyday language—is to some extent vague. . . . Hence . . . every solution of this problem implies necessarily a certain deviation from the practice of everyday language.
>
> In spite of all this, I happen to believe that the semantic conception does conform to a considerable extent with the common-sense usage . . . (*ibid.*, p. 360).

In setting up his problem, Tarski does not distance himself from the project of characterizing concepts that can be used as the ordinary semantic concepts are used; concepts that express, as he says, "connexions between the expressions of a language and the objects and states of affairs referred to by these expressions."[28] He does not aim, he says, to assign a new meaning to an old word, but to "catch hold of the actual meaning of an old notion."[29] In other words, he is quite explicit that he did not, as Etchemendy maintains, intend his definitions to be purely stipulative.

Tarski describes his project as "The Establishment of Scientific Semantics," and he says that "semantic concepts express certain relations between objects (and states of affairs) referred to in the language discussed and expressions of the language referring to these objects."[30] He regards the truth of a sentence as its "correspondence with reality" (*ibid.*). Tarski regards these characterizations of semantic concepts as "vague," but clearly they would be totally wrong if semantic concepts had no empirical application. When Tarski requires that his definitions be "materially adequate and in accordance with ordinary usage," he argues that convention-T is just what assures us that the condition is met. The argument is this: given a language we understand, an interpreted language such as

[27] *Philosophy and Phenomenological Research*, IV (1944): 341–375.
[28] "The Establishment of Scientific Semantics," in *Logic, Semantics, Metamathematics* (New York: Oxford, 1956), p. 401.
[29] "The Semantic Conception of Truth," p. 341.
[30] "The Establishment of Scientific Semantics," pp. 403–4.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

English, we recognize as true all sentences of the form " 'Snow is white' is true if and only if snow is white." Tarski calls such sentences "partial definitions" of truth. Obviously, a definition that entails all such sentences will have the same extension as the intuitive concept of truth with which we started. To admit this is to count T-sentences as having empirical content; otherwise convention-T would have no point, nor would Tarski's insistence that he is interested in defining truth only for interpreted languages.

We must conclude, I think, that, if Etchemendy, Soames, and Putnam are right, Tarski totally mistook his own aim and the nature of his accomplishment. Yet surprisingly little needs to be done to reconcile Tarski with Etchemendy. Etchemendy allows, of course, that "Tarski introduced precisely the mathematical techniques needed for an illuminating account of the semantic properties of certain simple languages," and "[g]etting from a Tarskian definition of truth to a substantive account of the semantic properties of the object language may involve as little as the reintroduction of a primitive notion of truth" (*op. cit.*, pp. 59–60). The trick is just to add to Tarski's definition of a truth predicate for a language L (say, 's is true$_L$') the remark that Tarski's predicate holds for all and only the true sentences of L. Here, of course, the word 'true' expresses the real-life, substantive, undefined concept we need for serious semantics. Let us call this remark the *truth axiom*.

The first thing to notice is that, if the language was consistent before we added the truth axiom, the truth axiom cannot make it inconsistent as long as we do not formally endow our new predicate with any properties. It can *have* all sorts of interesting properties and no formal harm will be done if the properties are not explicitly entered in the theory; and no informal harm will be done if the additional properties do not lead to contradiction.

Adding the truth axiom is, from a formal point of view, harmless; it is also pointless. For we can just as well regard Tarski's truth predicate 's is true$_L$' as having the properties of our real-life predicate 's is true in L', as long as those properties do not create inconsistencies. The objection to this thought is that we can no longer feel confident that, if we were to specify all the properties of the real-life predicate, inconsistencies might result; we do not know exactly what our truth predicate means. The "definition" of truth is no longer a purely stipulative definition.

Consider a formalized object language and a metalanguage exactly like those described by Tarski in sections 2 and 3 of "The Concept of Truth in Formalized Languages."[31] Now add to the metalanguage

[31] In *Logic, Semantics, Metamathematics*.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Tarski's definitions leading up to and including the definition of truth; but do not *call* them definitions, and think of them as employing empirically significant expressions suitable for describing the semantics of the object language (which has been interpreted as about the calculus of classes by Tarski). According to Etchemendy, the difference between this new system and Tarski's original is extreme: the new system correctly describes the semantics of the object language, while Tarski's system merely defines a predicate that cannot be used to assert anything, true or false, about any particular interpreted language. Tarski's definitions turn the entailed T-sentences into logical truths; the new system leaves them instructive remarks about the truth conditions of sentences. But this mighty change does not touch the formal system in any way; it is a change in how we describe the system, not in the system itself. If Tarski's system is consistent, so is the new one.

The entire issue turns, then, on how we regard definitions. Some definitions clearly are intended to introduce new words; others aim to express substantive truths of one sort or another. As we have seen, Tarski did not intend his definitions to foist a new meaning on an old term, but to "catch hold of the actual meaning of an old notion."[32]

We should now glance back at the theme, to be found not only in Etchemendy, but also in Putnam and Soames, that Tarski's truth definitions cannot have anything to do with the semantics or interpretation of actual languages because, given his definitions, the relevant theorems (e.g., the T-sentences) are logical truths. In fact, they are logical truths only on the assumption that Tarski's truth definitions are purely stipulative, that they tell us everything there is to know about the predicates he defines. There is no reason to accept this assumption. A simple analogy will make this clear. Suppose we offer as a definition of the predicate '*x* is a solar planet' the following: *x* is a solar planet if and only if *x* is just one of the following: Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, Neptune, Pluto. This entails the P-sentence 'Neptune is a solar planet'. Is this last a logical truth? One may as well say so if our definition is purely stipulative, otherwise not. The question whether it is purely stipulative is not one that can be answered by studying the formal system; it concerns the intentions of the person making the definition. If we were simply presented with the defining sentence, we could hardly

[32] Etchemendy suggests that the 'if and only if' of a definition does not have the same meaning as the 'if and only if' of a substantive claim, but I do not think this remark can be taken seriously since the difference makes no difference at all within the system, and if we were to mark the supposed difference by introducing different symbols, the rules of inference in the system would have to be altered. Etchemendy says his suggestion was not intended to be taken seriously (private conversation).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

fail to notice that, if we interpret the words in more or less the usual way, it expresses a substantive truth. By appeal to convention-T, Tarski invites us to notice an analogous feature of his truth definitions.

What should we conclude about how Tarski intended us to take his definitions? The indications may seem ambiguous. On the one hand, we have his repeated and explicit claim that he wished to, and thought he had, "caught the actual meaning" of the intuitive concept of truth, so far as this was possible; on the other hand, he clearly depended on the fact that his definitions allowed the elimination of all explicitly semantic vocabulary to guarantee that his concept would not introduce inconsistencies into an otherwise consistent language. But does this show that Tarski was confused? I think not. Here is a way of viewing the matter.

Tarski's definitions endow his truth predicates with properties that ensure that they define the class of true sentences in a language. If the predicates have no further properties, we know they will not breed inconsistencies. This makes the predicates useful for certain purposes. If we think of the truth predicates as having further unspecified properties, we cannot be sure those properties will not make trouble if they are made explicit. But there is nothing to prevent us from working within Tarski's systems while acknowledging that the truth predicates may have further essential properties, as long as we make no use of the unspecified properties. In this way, we can take full advantage of Tarski's technical achievement while not treating the contents of his theories as "empty" or "merely" formal.

To view Tarski's work in this light is to admit that there is a sense in which he did not define the concept of truth, even for particular languages. He defined the class of true sentences by giving the extension of the truth predicate, but he did not give the meaning. This follows the moment we decide that T-sentences have empirical content, for this implies that there is more to the concept of truth than Tarski's definition tells us. My contention is not that Tarski may after all have captured a substantial concept of truth, but that we are not necessarily confused if we interpret his formal systems as empirical theories about languages. By doing so, we avoid two potentially crippling theses about truth, theses which, as we have seen, are fairly common today. One is that Tarski's work is largely unrelated to the concept of truth as we ordinarily understand it, so that, if we want to study the semantics of interpreted languages, we must take another tack. Charybdis is the thesis that, although Tarski's version of truth is merely disquotational, it says all there is to say about the concept of truth.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

My own view is that Tarski has told us much of what we want to know about the concept of truth, and that there must be more. There must be more because there is no indication in Tarski's formal work of what it is that his various truth predicates have in common, and this must be part of the content of the concept. It is not enough to point to convention-T as that indication, for it does not speak to the question how we know that a theory of truth for a language is correct. The concept of truth has essential connections with the concepts of belief and meaning, but these connections are untouched by Tarski's work. It is here that we should expect to uncover what we miss in Tarski's characterizations of truth predicates.

What Tarski has done for us is show in detail how to describe the kind of pattern truth must make, whether in language or in thought. What we need to do now is to say how to identify the presence of such a pattern or structure in the behavior of people.

II

If there were no more to be known about the concept of truth than we can learn from Tarski's definitions of truth predicates, we would have no clear use for the concept of truth aside from the minor convenience of its disquotational function, since Tarski has shown how to eliminate such predicates without semantic residue. Any connection of truth with meaning or belief would be moot. If we think of Tarski's definitions as purely stipulative, the theorems such predicates allow us to prove, in particular the T-sentences, are equivalent to truths of logic; unless we read more into the truth predicates than the definitions provide, these theorems cannot, therefore, yield empirical truths about the sentences of any language, and cannot be taken to give the truth conditions of sentences.

Tarski never claimed his truth predicates did more than pick out the class of true sentences in particular languages. He certainly did not think he had defined a general truth predicate, nor did he aim to exceed the limits of extensionality. Capturing meaning, as distinguished from extension, was no part of his project. Nor did it matter to him that there might be other ways of characterizing the same classes of sentences—ways that could be more illuminating for purposes other than his.

The two points are related, since there is no evident way of giving a general characterization of truth without introducing criteria quite different in kind from those to which Tarski appealed. It is sometimes suggested by advocates of a deflationary view of truth that convention-T provides an adequate answer to the question what Tarski's various truth predicates have in common. But we should not be satisfied with this idea. For in those cases where the object lan-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

guage is contained in the metalanguage, the requirement is merely syntactical: it tells us something about the predicates, but not much about the concept. In other cases its application depends on our prior understanding of the notion of translation, a concept far more obscure than that of truth. The central point is this: aside from our grasp of the concept of translation, convention-T gives us no idea how to tell in general when one of Tarski's truth predicates applies to a particular language. He does not define the concept of translation.[33]

We still lack, then, a satisfactory account of the general feature or features of the concept of truth which we cannot find in Tarski. Nevertheless, we can learn a great deal from Tarski. His constructions make it evident, for example, that, for a language with anything like the expressive power of a natural language, the class of true sentences cannot be characterized without introducing a relation like satisfaction, which connects words (singular terms, predicates) with objects. If we think of satisfaction as a generalized form of reference, Tarski has shown how the truth of sentences depends on the semantic features (i.e., reference) of certain proper parts of sentences. (Of course, Tarski no more defined the general concept of reference than that of truth.) Thus, even without an answer to the question of how we know when a definition of truth applies to a given language, Tarski has shown how the concept of truth can be used to give a clear description of a language. Of course, to give such a description, we must have a grasp of the concept of truth first; but we can have such a grasp without being able to formulate a systematic description of a language. Convention-T connects our untutored grasp of the concept with Tarski's ingenious machinery; it persuades us that the workings of the machinery accord with the concept as we knew it.

This, then, is what we can learn about the concept of truth from Tarski: since it is obvious that he has not defined the general concept of truth, we can ignore the suggestion that his stipulative definitions capture all there is to that concept. But there is no reason not to make use of the structure that went into Tarski's definitions. To do this, we do not need to make any change in Tarski's formal systems; once we realize that those systems do not reflect important aspects of

[33] Michael Williams says a deflationist thinks that "what carries over from language to language . . . is the utility, for each language, of having its own disquotation device"—"Scepticism and Charity," *Ratio* (New Series), 1 (1988), p. 180. But aside from the difficulty of assigning a clear meaning to the "utility" of a device, there is the fact that in one language we can talk of truth in another language; and here the generalization hinted at by Williams can do no better than convention-T, with its essential appeal to translation.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the concepts of truth and reference, we can think of the truth and reference (satisfaction) predicates as *primitives* in the clauses that go into Tarski's recursive characterizations of reference and truth. If we find that the word 'definition' sorts ill with the idea that predicates are primitives, we can drop the word; this will not change the system. But to honor the recognition that the semantic predicates are primitives, we can drop the final step that for Tarski turns recursive characterizations into explicit definitions, and view the results as axiomatized theories of truth.[34]

An axiomatized theory of truth may be compared with, say, Kolmogorov's axiomatization of probability, which puts clear constraints on the concept of probability, but leaves open such questions as whether probability is to be further characterized as relative frequency, degree of belief, or something else. Ramsey's axiomatic treatment of preference in the face of uncertainty, when applied to a particular agent, is analogous to an axiomatized theory of truth in the further respect that it yields a separate theory for each agent, just as Tarskian truth theories are peculiar to a language, or, as I shall go on to propose, to an individual.

Just as a Tarskian theory does not tell us how to determine that the theory applies to a particular language or speaker, so nothing in Ramsey's theories tells us when such a theory applies to a particular agent. The issue in the case of decision theory is in part to specify the conditions an agent must satisfy in order to be said to prefer one object or course of action to another. In the case of a theory of truth, what we want to know is how to tell when T-sentences (and hence the theory as a whole) describe the language of a group or an individual. This obviously requires specifying at least part of the content of the concept of truth which Tarski's truth predicates fail to capture.

What do we add, then, to the properties of truth that Tarski has

[34] Tarski recognized the possibility of giving axiomatic theories of truth, and remarked that "[t]here is nothing essentially wrong in such an axiomatic procedure, and it may prove useful for various purposes"—"The Semantic Conception of Truth," p. 352. Tarski had a number of reasons for preferring an explicit definition to an axiomatic treatment of the concept of truth. First, he notes that the choice of axioms "has a rather accidental character, depending on inessential factors (such as e.g. the actual state of our knowledge)." Second, only an explicit definition can guarantee the consistency of the resulting system (given the consistency of the system prior to introducing the new primitive concepts); and, third, only an explicit definition can subdue doubts whether the concept is "in harmony with the postulates of the unity of science and of physicalism"—"The Establishment of Scientific Semantics," pp. 405–6. The first danger is avoided if the axioms are restricted to the recursive clauses needed to characterize satisfaction; the second is (less conclusively) evaded as long as known ways of producing paradox are not introduced; and the threat that truth might not turn out to be reducible to physical concepts is a threat that, in my opinion, we neither can nor should want to escape.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

delineated when we apply the intuitive concept of truth? Aside from the view that Tarski said all that can or should be said about truth, a view I discussed and rejected in the first section of this essay, I think most contemporary proposals fall into two broad categories: those which humanize truth by making it basically epistemic, and those which promote some form of correspondence theory.

Many philosophers, particularly recently, have held that truth is an epistemic concept; even when they have not explicitly held this thesis, their views have often implied it. Coherence theories of truth are usually driven by an epistemic engine, as are pragmatic characterizations of truth. The antirealism of Dummett and Crispin Wright, Peirce's idea that the truth is where science will end up if it continues long enough, Richard Boyd's claim that truth is what explains the convergence of scientific theories, and Putnam's internal realism all include or entail an epistemic account of truth. Quine also, at least at times, has maintained that truth is internal to a theory of the world and so to that extent is dependent on our epistemological stance. Relativism about truth is perhaps always a symptom of infection by the epistemological virus; this seems to be true in any case for Quine, Nelson Goodman, and Putnam.

Apparently opposed to these views is the intuitive idea that truth, aside from a few special cases, is entirely independent of our beliefs; as it is sometimes put, our beliefs might be just as they are and yet reality—and so the truth about reality—be very different. According to this intuition, truth is "radically non-epistemic" (so Putnam characterizes "transcendental realism"), or "evidence-transcendent" (to use Dummett's phrase for realism). (Both Putnam and Dummett are, of course, opposed to such views.) If we were to look for tags for these two views of truth, we might hit on the adjectives 'epistemic' and 'realist'; the assertion of an essential tie to epistemology introduces a dependence of truth on what can somehow be verified by finite rational creatures, while the denial of any dependence of truth on belief or other human attitudes defines one philosophical use of the word 'realism'.

In the next, and last, section of this essay, I outline an approach to the concept of truth which rejects both of these views of truth. I do not aim to reconcile the two positions. I find epistemic views untenable, and realist views ultimately unintelligible. That both views, while no doubt answering to powerful intuitions, are fundamentally mistaken is at least suggested by the fact that both invite skepticism. Epistemic theories are skeptical in the way idealism or phenomenalism are skeptical; they are skeptical not because they make reality unknowable, but because they reduce reality to so much less than we

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

believe there is. Realist theories, on the other hand, seem to throw in doubt not only our knowledge of what is "evidence-transcendent," but all the rest of what we think we know, for such theories deny that what is true is conceptually connected in any way to what we believe.

Let us consider the project of giving content to a theory of truth. Tarski's definitions are normally reached through several steps. First, there is a definition of what it is to be a sentence in the object language; then a recursive characterization of a satisfaction relation (satisfaction is a highly generalized version of reference); the recursive characterization of satisfaction is turned into an explicit definition in the manner of Gottlob Frege and Dedekind; then truth is defined on the basis of the concepts of sentence and satisfaction. We are dropping the step that turns the recursive characterization of satisfaction into a definition, thus making explicit the fact that we are treating the truth and satisfaction predicates as primitives.

Which of the two semantic concepts, satisfaction or truth, we take as basic is, from a formal point of view, open to choice. Truth, as Tarski showed, is easily defined on the basis of satisfaction; but, alternatively, satisfaction can be taken to be whatever relation yields a correct account of truth. Tarski's work may seem to give uncertain signals. The fact that the truth of sentences is defined by appeal to the semantic properties of words suggests that, if we could give a satisfactory account of the semantic properties of words (essentially, of reference or satisfaction), we would understand the concept of truth. On the other hand, the key role of convention-T in determining that truth, as characterized by the theory, has the same extension as the intuitive concept of truth, makes it seem that it is truth rather than reference that is the basic primitive. The second is, I think, the right view. In his appeal to convention-T, Tarski assumes, as we have seen, a prior grasp of the concept of truth; he then shows how this intuition can be implemented in detail for particular languages. The implementation requires the introduction of a referential concept, a relation between words and things—some relation like satisfaction. The story about truth generates a pattern in language, the pattern of logical forms, or grammar properly conceived, and the network of semantic dependencies. There is no way to tell this story, which, being about truth, is about sentences or their occasions of use, without assigning semantical roles to the parts of sentences. But there is no appeal to a prior understanding of the concept of reference.

This way of viewing a theory of truth runs contrary to a tradition. According to the tradition, we could never come to understand sentences in their vast or even infinite array unless we understood the words drawn from a finite vocabulary which make them up; there-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

fore, the semantic properties of words must be learned before we understand sentences, and the semantic properties of words have conceptual priority because it is they that explain the semantic properties—above all the truth conditions—of sentences. I think this line of argument, which starts with a truism, ends with a false conclusion; so something must go wrong. The mistake is to confuse the order of explanation that is appropriate once the theory is in place with the explanation of why the theory is correct. The theory is correct because it yields correct T-sentences; its correctness is tested against our grasp of the concept of truth as applied to sentences. Since T-sentences say nothing whatever about reference, satisfaction, or expressions that are not sentences, the test of the correctness of the theory is independent of intuitions concerning these concepts. Once we have the theory, though, we can explain the truth of sentences on the basis of their structure and the semantic properties of the parts. The analogy with theories in science is complete: in order to organize and explain what we directly observe, we posit unobserved or indirectly observed objects and forces; the theory is tested by what is directly observed.

The perspective on language and truth that we have gained is this: what is open to observation is the use of sentences in context, and truth is the semantic concept we understand best. Reference and related semantic notions like satisfaction are, by comparison, theoretical concepts (as are the notions of singular term, predicate, sentential connective, and the rest). There can be no question about the correctness of these theoretical concepts beyond the question whether they yield a satisfactory account of the use of sentences.

One effect of these reflections is to focus on the centrality of the concept of truth in the understanding of language; it is our grasp of this concept that permits us to make sense of the question whether a theory of truth for a language is correct. There is no reason to look for a prior, or independent, account of some referential relation. The other main consequence of the present stance is that it provides an opportunity to say fairly sharply what is missing, as an account of truth, in a theory of truth in Tarski's style.

What is missing is the connection with the users of language. Nothing would count as a sentence, and the concept of truth would therefore have no application, if there were not creatures who used sentences by uttering or inscribing tokens of them. Any complete account of the concept of truth must relate it to actual linguistic intercourse. Put more precisely: the question whether a theory of truth is true of a given language (that is, of a speaker or group of speakers) makes sense only if the sentences of that language have a

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

meaning that is independent of the theory (otherwise the theory is not a theory in the ordinary sense, but a description of a possible language). Or to return to the definitional form that Tarski favored: if the question can be raised whether a truth definition really does define truth for a given language, the language must have a life independent of the definition (otherwise the definition is merely stipulative: it specifies, but is not true of, a language).

If we knew in general what makes a theory of truth correctly apply to a speaker or group of speakers, we could plausibly be said to understand the concept of truth; and if we could say exactly what makes such a theory true, we could give an explicit account—perhaps a definition—of truth. The ultimate evidence, as opposed to a criterion, for the correctness of a theory of truth must lie in available facts about how speakers use the language. When I say available, I mean publicly available—available not only in principle, but available in practice to anyone who is capable of understanding the speaker or speakers of the language. Since all of us do understand some speakers of some languages, all of us must have adequate evidence for attributing truth conditions to the utterances of some speakers; all of us have, therefore, a competent grasp of the concept of truth as applied to the speech behavior of others.

Have we now settled the question whether truth is radically nonepistemic, as realists aver, or basically epistemic, as others maintain? It may seem that the matter is settled in favor of the epistemic or subjective view, since we have followed a course of argument that leads to the conclusion that it is how language is used which decides whether a theory of truth for that language is true. But in fact the matter is not settled, for it may be held by realists that the question whether the theory is true for a given language or group of speakers is indeed empirical, but only because the question of what the words mean is empirical; the issue of truth, it may be held, remains to be answered, whether by the theory itself or in some other way.

Does the theory already contain the answer? It does if there is substance to the claim that a Tarski-type theory of truth is a correspondence theory, for then the theory must in effect define truth to be correspondence with reality—the classical form of realism with respect to truth. Tarski himself said he wanted his truth definitions to "do justice to the intuitions which adhere to the *classical conception of truth*"; he then quotes Aristotle's *Metaphysics* Γ ("to say of what is that it is, or of what is not that it is not, is true"), and offers as an alternative formulation,

> *The truth of a sentence consists in its agreement with (or correspondence to) reality.*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(Tarski[35] adds that the term 'correspondence theory' has been suggested for this way of putting things.) I have myself argued in the past that theories of the sort Tarski showed how to produce were correspondence theories of a sort.[36] I said this on the ground that there is no way to give such a theory without employing a concept like reference or satisfaction which relates expressions to objects in the world.

It now seems to me to have been a mistake to call such theories correspondence theories. Here is the reason I think it was a mistake. The usual complaint about correspondence theories is that it makes no sense to suggest that it is somehow possible to compare one's words or beliefs with the world, since the attempt must always end up simply with the acquisition of more beliefs. This complaint was voiced, for example, by Otto Neurath,[37] who for this reason adopted a coherence view of truth; Carl Hempel[38] has expressed the same objection, speaking of the "fatal confrontation of statements and facts" (*ibid.*, p. 51). Rorty[39] has repeatedly insisted, claiming sympathy with Dewey, that a correspondence view of truth makes the concept of truth useless. I have said much the same myself.[40]

This complaint against correspondence theories is not sound. One reason it is not sound is that it depends on assuming that some form of epistemic theory is correct; therefore, it would be a legitimate

[35] "The Semantic Conception of Truth," pp. 342–3. Tarski also speaks of sentences "describing" "states of affairs," *ibid.*, p. 345. Cf. "The Concept of Truth in Formalized Languages," p. 153, and "The Establishment of Scientific Semantics," p. 403.

[36] In "True to the Facts," in *Inquiries into Truth and Interpretation*. The argument is this. Truth is defined on the basis of satisfaction: a sentence of the object language is true if and only if it is satisfied by every sequence of the objects over which the variables of quantification of the object language range. Take 'corresponds to' as "satisfies" and you have defined truth as correspondence. The oddity of the idea is evident from the counterintuitive and contrived nature of the entities to which sentences "correspond" and from the fact that all true sentences would correspond to the same entities.

[37] "Protokollsätze," *Erkenntnis*, III (1932/33): 204–214.

[38] "On the Logical Positivist's Theory of Truth," *Analysis*, II (1935): 49–59.

[39] *Consequences of Pragmatism*, "Introduction"; also "Pragmatism, Davidson and Truth," in Ernest LePore, ed., *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (New York: Blackwell, 1986).

[40] The position I take in the present paper was influenced by an exchange between Rorty and me at the 1982 meeting of the Pacific Division of the American Philosophical Association. Rorty persuaded me not to call my position either a correspondence theory or a coherence theory; I think I persuaded him to give up the pragmatist theory of truth. "Pragmatism, Davidson and Truth" is a revised version of Rorty's 1982 talk at the Pacific Division meeting.

For an example of the use of 'correspondence' I now deplore, see my "A Coherence Theory of Truth and Knowledge," in *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

complaint only if truth were an epistemic concept. If this were the only reason for rejecting correspondence theories, the realist could simply reply that his position is untouched; he always maintained that truth was independent of our beliefs or our ability to learn the truth.

The real objection to correspondence theories is simpler; it is that there is nothing interesting or instructive to which true sentences might correspond. The point was made some time ago by C. I. Lewis;[41] he challenged the correspondence theorist to *locate* the fact or part of reality, or of the world, to which a true sentence corresponded. One can locate individual objects, if the sentence happens to name or describe them, but even such location makes sense relative only to a frame of reference, and so presumably the frame of reference must be included in whatever it is to which a true sentence corresponds. Following out this line of thought led Lewis to conclude that, if true sentences correspond to anything at all, it must be the universe as a whole; thus, all true sentences correspond to the same thing. Frege, as we know, reached the same conclusion through a somewhat similar course of reasoning. Frege's argument, if Alonzo Church[42] is right, can be formalized: starting from the assumptions that a true sentence cannot be made to correspond to something different by the substitution of coreferring singular terms, or by the substitution of logically equivalent sentences, it is easy to show that, if true sentences correspond to anything, they all correspond to the same thing. But this is to trivialize the concept of correspondence completely; there is no interest in the relation of correspondence if there is only one thing to which to correspond, since, as in any such case, the relation may as well be collapsed into a simple property: thus, '*s* corresponds to the universe', like '*s* corresponds to (or names) the True', or '*s* corresponds to the facts' can less misleadingly be read '*s* is true'. Peter Strawson[43] has observed that the parts of a sentence may correspond to parts of the world (that is, refer to them), but adds,

> It is evident that there is nothing else in the world for the statement itself to be related to. . . . And it is evident that the demand that there should be such a relatum is logically absurd. . . . But the demand for something in the world *which makes the statement true* . . . , or *to which the*

[41] *An Analysis of Knowledge and Valuation* (La Salle, IL: Open Court, 1946), pp. 50–55.

[42] The argument, attributed to Frege by Church, can be found in Church's *Introduction to Mathematical Logic, Vol. 1* (Princeton: University Press, 1956), pp. 24–5. Frege's argument is rehearsed in my "True to the Facts."

[43] "Truth," in *Logico-Linguistic Papers* (London: Methuen, 1971).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

statement corresponds when it is true, is just this demand (*ibid.*, pp. 194–5).

He correctly goes on to claim that, "while we certainly say that a statement corresponds to (fits, is borne out by, agrees with) the facts," this is merely "a variant on saying it is true" (*ibid.*).

The correct objection to correspondence theories is not, then, that they make truth something to which humans can never legitimately aspire; the real objection is rather that such theories fail to provide entities to which truth vehicles (whether we take these to be statements, sentences or utterances) can be said to correspond. If this is right, and I am convinced it is, we ought also to question the popular assumption that sentences, or their spoken tokens, or sentence-like entities or configurations in our brains, can properly be called "representations," since there is nothing for them to represent. If we give up facts as entities that make sentences true, we ought to give up representations at the same time, for the legitimacy of each depends on the legitimacy of the other.

There is thus a serious reason to regret having said that a Tarski-style truth theory was a form of correspondence theory. My basic reason for saying it was not that I had made the mistake of supposing sentences or utterances of sentences corresponded to anything in an interesting way. But I was still under the influence of the idea that there is something important in the realist conception of truth; the idea that truth, and therefore reality, are (except for special cases) independent of what anyone believes or can know. Thus, I advertised my view as a brand of realism, realism with respect to the "external world," with respect to meaning, and with respect to truth.[44]

The terms 'realism' and 'correspondence' were ill-chosen because they suggest the positive endorsement of a position, or an assumption that there is a clear positive thesis to be adopted, whereas all I was entitled to maintain, and all that my position actually entailed with respect to realism and truth, was the negative view that epistemic views are false. The realist view of truth, if it has any content, must be based on the idea of correspondence, correspondence as applied to sentences or beliefs or utterances—entities that are propositional in character; and such correspondence cannot be made intelligible. I simply made the mistake of assuming realism and epistemic theories were the only possible positions. The only legitimate reason I had for calling my position a form of realism was to reject positions like Dummett's antirealism; I was concerned to reject the

[44] "A Coherence Theory of Truth and Knowledge," p. 307.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

doctrine that either reality or truth depends directly on our epistemic powers. There is a point in such a rejection. But it is futile either to reject or to accept the slogan that the real and the true are "independent of our beliefs." The only evident positive sense we can make of this phrase, the only use that consorts with the intentions of those who prize it, derives from the idea of correspondence, and this is an idea without content.[45]

To reject the doctrine that the real and true are independent of our beliefs is not, of course, to reject the platitude it can mistakenly be thought to express: believing something does not in general make it true. For allowing that the platitude is true does not commit us to saying there is no connection whatever between belief and truth; there must be some connection if we are to relate the truth of utterances to their use. The question is what that connection can be.

Various forms of subjectivism—that is, of views that make truth out to be an epistemic concept—connect human thoughts, desires, and intentions to truth in quite different ways, and I cannot pretend to do justice to all such views here. The best I can do is indicate why, despite the differences among the various positions, it makes sense to be dissatisfied with them all.

I have classified coherence theories of truth as epistemic, and this needs an explanation. A pure coherence theory of truth would hold, I suppose, that all the sentences in a consistent set of sentences are true. Perhaps no one has ever held such a theory, for it is mad. Those who have proposed coherence theories, for example Neurath and Rudolf Carnap (at one time), have usually made clear that it was sets of *beliefs*, or of sentences held to be true, whose consistency was enough to make them true; this is why I class coherence theories with epistemic views: they tie truth directly to what is believed. But unless something further is added, this view seems as wrong as Moritz Schlick[46] held it to be (he called it an "astounding error"); the obvious objection is that many different consistent sets of beliefs are possible which are not consistent with one another.[47]

[45] Arthur Fine gives up on realism for some of the same reasons I do, and he adds a splendid refutation of the thesis that a realistic view of truth explains the practice and advance of science—"The Natural Ontological Attitude," in *The Shaky Game: Einstein, Realism and the Quantum Theory* (Chicago: University Press, 1986).

[46] "Über das Fundament der Erkenntnis," *Erkenntnis,* IV (1934): 79–99.

[47] Not every theory that relates truth to coherent sets of beliefs is wrong. What must be added to standard coherence theories is an appreciation not only of how beliefs are causally and logically related to each other, but of how the contents of a belief depends on its causal connections with the world. I discuss these matters in the next section. Also see my "A Coherence Theory of Truth and Knowledge" and "Empirical Content," in *Truth and Interpretation: Perspectives on the Philosophy*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

There are theories, similar in certain ways to coherence theories, which have much the same drawback. Quine holds that the truth of some sentences, which he calls observation sentences, is tied directly to experience (more precisely, to patterns of excited nerve endings); further sentences derive their empirical content from their connections with observation sentences and their logical relations to one another. The truth of the resulting theory depends only on how well it serves to explain or predict true observation sentences. Quine plausibly maintains that there could be two theories equally capable of accounting for all true observation sentences, and yet such that neither theory can be reduced to the other (each theory contains at least one predicate that cannot be defined using the resources of the other theory). Quine has at different times embraced different ways of thinking of this situation. According to one way, both theories are true. I see no reason to object to the view that empirically equivalent theories (however one characterizes empirical content) are true or false together. According to Quine's other view, a speaker or thinker at a given time operates with one theory and, for him at that time, the theory he is using is true and the other theory false. If he shifts to the alternative theory, then *it* becomes true and the previously accepted theory false. The position may illustrate what Quine means when he says that truth is "immanent."[48] This conception of the immanence or relativity of truth should not be confused with the pedestrian sense in which the truth of sentences is relative to the language in which they occur. Quine's two theories can belong to, and be stated in, the same language; indeed, they must be if we are to understand the claim that the theories conflict. It is not easy to see how the same sentence (without indexical elements), with interpretation unchanged, can be true for one person and not for another, or for a given person at one time and not at another. The difficulty seems due to the attempt to import epistemological considerations into the concept of truth.

---

of Donald Davidson. It now seems to me one more terminological error to have called the thesis of "A Coherence Theory" a coherence theory. I explain why at more length in "Afterthoughts, 1987," added to a reprinting of "A Coherence Theory" to appear in A. Malichowski, ed., *Reading Rorty* (New York: Blackwell, 1990), pp. 136–8.

[48] See *Ontological Relativity and Other Essays* (New York: Columbia, 1969). For Quine's problem about empirically equivalent, mutually irreducible theories, see his "On Empirically Equivalent Systems of the World," *Erkenntnis*, IX (1975): 313–328; *Theories and Things* (Cambridge: Harvard, 1981), pp. 29–30; L. E. Hahn and P. A. Schilpp, eds., *The Philosophy of W. V. Quine* (La Salle, IL: Open Court, 1986), pp. 156–7.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Putnam's "internal realism" also makes truth immanent, though not, as Quine's view does, relative to a theory, but to the entire language and conceptual scheme a person accepts. Of course, if all this means is that the truth of sentences or utterances is relative to a language, that is familiar and trivially correct. But Putnam seems to have something more in mind—for example, that a sentence of yours and a sentence of mine may contradict each other, and yet each be true "for the speaker." It is hard to think in what language this position can be coherently, much less persuasively, expressed. The source of the trouble is once again the felt need to make truth accessible. Putnam is clear that this is the consideration that concerns him. He identifies truth explicitly with idealized justified assertability. He calls this a form of realism because there is "a fact of the matter as to what the verdict would be if the conditions were sufficiently good, a verdict to which opinion would 'converge' if we were reasonable."[49] He adds that his view is "a *human* kind of realism, a belief that there is a fact of the matter as to what is rightly assertable for us, as opposed to what is rightly assertable from the God's eye view so dear to the classical metaphysical realist" (*ibid.*). One suspects that, if the conditions under which someone is ideally justified in asserting something were spelled out, it would be apparent either that those conditions allow the possibility of error or that they are so ideal as to make no use of the intended connection with human abilities. It is also striking that Putnam seems to have no argument for his position except that the alternative ("metaphysical realism"—i.e., a correspondence theory) is unacceptable. He does not argue that there can be no other position.

Putnam describes his position as close to Dummett's on the main point—the epistemological status of truth. One difference is that Putnam is less certain than Dummett that truth is limited to what is definitely ascertainable, and therefore he is less sure that the principle of bivalence must be abandoned; this perhaps explains why Putnam calls his view a form of realism while Dummett calls his position antirealist. Putnam also thinks he differs from Dummett in tying truth to *idealized* justified assertability instead of justified assertability; but here I think a close reading of Dummett would show that he has much the same idea. If Dummett does not insist on something similar to Putnam's ideal conditions, then I think a criticism of Dummett that Putnam once formulated applies: if truth depends simply on justified assertability, truth can be "lost," that is, a sen-

[49] *Realism and Reason: Philosophical Papers, Volume 3* (New York: Cambridge, 1983), p. xviii.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tence can be true for a person at a time and later become false because the conditions of justification change. This must be wrong.[50] Dummett says he agrees that truth cannot be lost, but he fails to give a clear idea of how warranted assertability can be both a fixed property *and* a property that depends on the actual ability of human speakers to recognize that certain conditions are satisfied. Actual abilities wax and wane, and differ from person to person; truth does not.

Why does Dummett endorse this view of truth? There are a number of reasons, but one seems to be this. We have seen that a theory of truth in Tarski's style neither defines nor fully characterizes truth; there is no way to tell if the theory applies to a speaker or group of speakers unless something is added to relate the theory to the human uses of language. Dummett thinks the only way to do this is to make truth humanly recognizable. The human use of language must be a function of how people understand the language, so if truth is to play a role in explaining what it is to understand a language, there must be something, Dummett thinks, that counts as a person having "conclusive evidence" that a statement is true. One can appreciate the force of this idea while finding it difficult to accept. I have given my chief reason for rejecting it; that it is either empty, or makes truth a property that can be lost. But it is important to realize that there are other strong intuitions that would also have to be sacrificed if Dummett were right. One is the connection of truth with meaning: on Dummett's view, we can understand a sentence like 'A city will never be built on this spot' without having any idea what it would be for this sentence to be true (since the sentence, or an utterance of it, has no truth value for Dummett). Another is the connection of truth with belief: on Dummett's view, I can understand and believe a city will never be built on this spot, but my belief will have no truth value. It would seem that, for Dummett, having a belief that one expresses by a given sentence is not necessarily to believe that the sentence is true.

I might be tempted to go along with Dummett if I thought we must choose between what Putnam calls transcendental realism, i.e., the view that truth is "radically non-epistemic," that all our best researched and established beliefs and theories may be false, and Dummett's identification of truth with warranted assertability, since I find the former view—essentially the correspondence view—in-

[50] Putnam, "Reference and Understanding" and "Reply to Dummett's Comment," in A. Margalit, ed., *Meaning and Use* (Dordrecht: Reidel, 1979), pp. 226–8.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

comprehensible, while I find Dummett's view merely false. But I see no reason to suppose that realism and antirealism, explained in terms of the radically nonepistemic or the radically epistemic character of truth, are the only ways to give substance to a theory of truth or meaning.

Let us briefly take stock. In the first section of this essay, I rejected deflationary views of truth, those which teach that there is no more to the concept than Tarski has shown how to define for particular languages. In this section, I have argued that certain familiar attempts to characterize truth which go beyond giving empirical content to a structure of the sort Tarski taught us how to describe are empty, false, or confused. We should not say that truth is correspondence, coherence, warranted assertability, ideally justified assertability, what is accepted in the conversation of the right people, what science will end up maintaining, what explains the convergence on single theories in science, or the success of our ordinary beliefs. To the extent that realism and antirealism depend on one or another of these views of truth we should refuse to endorse either. Realism, with its insistence on radically nonepistemic correspondence, asks more of truth than we can understand; antirealism, with its limitation of truth to what can be ascertained, deprives truth of its role as an intersubjective standard. We must find another way of viewing the matter.

III

A theory of truth, in contrast to a stipulative definition of truth, is an empirical theory about the truth conditions of every sentence in some corpus of sentences. But of course sentences are abstract objects, shapes, say, and do not have truth conditions except as embodied in sounds and scribbles by speakers and scribblers. In the end it is the utterances and writings of language users with which a theory of truth must deal; the role of sentences in a theory is merely to make it possible to deal with *types* of utterances and inscriptions, whether or not particular types are realized. Introducing sentences thus serves two purposes: it allows us to speak of all actual utterances and inscriptions of the same type in one breath; and it allows us to stipulate what the truth conditions of an utterance or inscription of a given type would be if it were uttered. (For brevity I shall from here on refer to acts of writing as utterances along with their audible counterparts.)

Although we may sometimes say a group speaks with one voice, utterances are essentially personal; each utterance has its agent and its time. An utterance is an event of a special sort, an intentional action. Theories of truth are primarily concerned with sentential

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

utterances, utterances which, whatever their surface grammar, must be treated as utterances of sentences. The primacy of sentences or sentential utterances is dictated by the fact that it is sentences, as uttered on particular occasions by particular speakers, for which the theory supplies truth conditions and of which truth is predicated. Verbal felicity apart, there is no reason not to call the utterance of a sentence, under conditions that make the sentence true, a true utterance.

A theory of truth does more than describe an aspect of the speech behavior of an agent, for it not only gives the truth conditions of the actual utterances of the agent; it also specifies the conditions under which the utterance of a sentence would be true if it were uttered. This applies both to sentences actually uttered, by telling us what would have been the case if those sentences had been uttered at other times or under other circumstances, and to sentences never uttered. The theory thus describes a certain complex ability.

An utterance has certain truth conditions only if the speaker intends it to be interpreted as having those truth conditions. Moral, social, or legal considerations may sometimes invite us to deny this, but I do not think the reasons for such exceptions reveal anything of importance about what is basic to communication. Someone may say something that would normally be offensive or insulting in a language he believes his hearers do not understand; but in this case his audience for the purpose of interpretation is obviously just the speaker himself. A malapropism or slip of the tongue, if it means anything, means what its promulgator intends it to mean. There are those who are pleased to hold that the meanings of words are magically independent of the speaker's intentions; for example, that they depend on how the majority, or the best-informed, or the best-born, of the community in which the speaker lives speak, or perhaps how they would speak if they took enough care.[51] This doctrine entails that a speaker may be perfectly intelligible to his hearers, may be interpreted exactly as he intends to be interpreted, and yet may not know what he means by what he says. I think this view, though it has been elaborately and ingeniously defended, reveals nothing of seri-

[51] Saul Kripke attributes such a view to Wittgenstein in *Wittgenstein on Rules and Private Language* (New York: Blackwell, 1982), and tentatively endorses it. For a different version, see the numerous works of Tyler Burge on anti-individualism, e.g., "Individualism and the Mental," in P. French, T. Uehling, H. Wettstein, eds., *Midwest Studies in Philosophy, Volume 4* (Minneapolis: Minnesota UP, 1979), pp. 73–121; "Individualism and Psychology," *Philosophical Review*, XCV (1986): 3–46; "Wherein is Language Social?" in A. George, ed., *Reflections on Chomsky* (New York: Blackwell, 1989), pp. 176–191.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ous philosophical interest about the nature of truth or meaning (though it may have much to do with good, or acceptable, manners, and may represent an intention, or even some sort of social responsibility, on the part of some speakers).[52] For the purpose of the present enterprise, that of understanding truth and meaning, we should, I think, stick as closely as possible to what is made directly available to an audience by a speaker, and this is the relevant state of the speaker's mind. What matters to successful linguistic communication is the intention of the speaker to be interpreted in a certain way, on the one hand, and the actual interpretation of the speaker's words along the intended lines through the interpreter's recognition of the speaker's intentions, on the other.[53]

The approach I am following puts no primary weight on the concept of a language as something shared by speaker and interpreter, or by a speaker and a speech community, except in this sense: though communication by speech does not, as far as I can see, require that any two speakers *speak* in the same way, it does, of course, demand a fit between how speakers intend to be interpreted and how their interpreters understand them. This demand no doubt tends to encourage convergence in speech behavior among those who exchange words, the degree depending on factors like shared social and economic status, educational and ethnic background, etc. What convergence exists is of such vast practical importance that we may exaggerate both its degree and its philosophical significance. But I think we do well to ignore this practical issue in constructing theories of meaning, of truth, and of linguistic communication.[54] I shall therefore treat theories of truth as applicable in the first place to individual speakers at various periods or even moments of their lives.

A theory of truth links speaker with interpreter: it at once describes the linguistic abilities and practices of the speaker and gives

[52] See my "Knowing One's Own Mind," *Proceedings and Addresses of the American Philosophical Association,* LX (1987): 441–458.

[53] The influence of H. P. Grice's "Meaning," *The Philosophical Review,* LXVI (1957): 377–388, will be evident here.

My characterization of successful communication leaves open a range of possibilities with respect to the question what a speaker means by her words on occasion. Since the speaker must intend to be interpreted in a certain way, she must believe her audience is equipped to interpret her words in that way. But how well justified must this belief be, and how nearly correct? I do not believe our standards for deciding what someone's words, as spoken on a given occasion, mean are firm enough to let us draw a sharp line between a failed intention that one's words have a certain meaning and a success at meaning accompanied by a failed intention to be interpreted as intended.

[54] See my "Communication and Convention," in *Inquiries into Truth and Interpretation.*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the substance of what a knowledgeable interpreter knows which enables him to grasp the meaning of the speaker's utterances. This is not to say that either speaker or interpreter is aware of or has propositional knowledge of the contents of such a theory. The theory describes the conditions under which an utterance of a speaker is true, and so *says* nothing directly about what the speaker knows. The theory does, however, *imply* something about the propositional content of certain intentions of the speaker, namely, the intentions that his utterances be interpreted in a certain way. And though the interpreter certainly does not need to have explicit knowledge of the theory, the theory does provide the only way to specify the infinity of things the interpreter knows about the speaker, namely, the conditions under which each of an indefinitely large number of sentences of the speaker would be true if uttered.

There must, of course, be *some* sense in which speaker and interpreter have internalized a theory; but this comes to no more than the fact that the speaker is able to speak as if he believed the interpreter would interpret him in the way the theory describes, and the fact that the interpreter is prepared so to interpret him. All we should require of a theory of truth for a speaker is that it be such that, *if* an interpreter had explicit propositional knowledge of the theory, he would know the truth conditions of utterances of the speaker.[55]

A theory of truth for a speaker is a theory of meaning in this sense, that explicit knowledge of the theory would suffice for understanding the utterances of that speaker. It accomplishes this by describing the critical core of the speaker's potential and actual linguistic behavior, in effect, how the speaker intends his utterances to be interpreted. The sort of understanding involved is restricted to what we may as well call the literal meaning of the words, by which I mean, roughly, the meaning the speaker intends the interpreter to grasp, whatever further force or significance the speaker wants the interpreter to fathom.[56]

[55] This is, of course, far more than is provided by any theory anyone has been able to provide for any natural language. The condition is therefore not one we know could be satisfied. We do, on the other hand, know how to produce such a theory for a powerful, perhaps self-sufficient, fragment of English and other natural languages, and this is enough to give substance to the idea that the incorporation of the concept of truth in a theory provides insight into the nature of the concept. We may in the end have to settle for a far looser sense of 'theory' than Tarski had in mind.

I am skipping a host of well-exercised problems, such as providing truth conditions for subjunctive conditionals, imperatives, interrogatives, ethical statements, etc. I have discussed (though certainly not solved) most of these problems elsewhere.

[56] There is one intention not touched on by a theory of truth which a speaker must intend an interpreter to perceive, the *force* of the utterance. An interpreter

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The thesis that a theory of truth conditions gives an adequate account of what is needed for understanding the literal meanings of utterances is, of course, much disputed, but since I have argued for it at length elsewhere, I will for the most part treat the thesis here as an assumption. If the assumption is mistaken, much of the detail in what I shall go on to say about the application of the concept of truth will be threatened, but the general approach will, I think, remain valid.

A theory of truth, viewed as an empirical theory, is tested by its relevant consequences, and these are the T-sentences entailed by the theory. A T-sentence says of a particular speaker that, every time he utters a given sentence, the utterance will be true if and only if certain conditions are satisfied. T-sentences thus have the form and function of natural laws; they are universally quantified bi-conditionals, and as such are understood to apply counterfactually and to be confirmed by their instances.[57] Thus, a theory of truth is a theory for describing, explaining, understanding, and predicting a basic aspect of verbal behavior. Since the concept of truth is central to the theory, we are justified in saying truth is a crucially important explanatory concept.

The question that remains is: How do we confirm the truth of a T-sentence? The question is a kind of question that arises with respect to many theories, both in the physical sciences and in psychology. A theory of fundamental measurement of weight, for example, states in axiomatic form the properties of the relation between x and y that holds when x is at least as heavy as y; this relation must, among other things, be transitive, reflexive, and nonsymmetric. A theory of preference may stipulate that the relation of weak preference has the same formal properties. But in neither case do the axioms define the central relation (x is at least as heavy as y, x is weakly preferred to y), nor instruct us how to determine when the relation holds. Before the

---

must, if he is to understand a speaker, be able to tell whether an utterance is intended as a joke, an assertion, an order, a question, and so forth. I do not believe there are rules or conventions that govern this essential aspect of language. It is something language users can convey to hearers and hearers can, often enough, detect; but this does not show that these abilities can be regimented. I think there are sound reasons for thinking nothing like a serious theory is possible concerning this dimension of language. Still less are there conventions or rules for creating or understanding metaphors, irony, humor, etc. See my "What Metaphors Mean" and "Convention and Communication," in *Inquiries into Truth and Interpretation*.

[57] This goes some way to answer a frequent criticism of theories of truth as theories of meaning. For example, given the (unusual) case of two unstructured predicates with the same extension, a theory of truth may make a distinction if there are circumstances which never arise but under which the truth conditions would be different.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

theory can be tested or used, something must be said about the interpretation of the undefined concepts. The same applies to the concept of truth.[58]

It is a mistake to look for a behavioristic definition, or indeed any other sort of explicit definition or outright reduction of the concept of truth. Truth is one of the clearest and most basic concepts we have, so it is fruitless to dream of eliminating it in favor of something simpler or more fundamental. Our procedure is rather this: we have asked what the formal properties of the concept are when it is applied to relatively well-understood structures, namely, languages. Here Tarski's work provides the inspiration. It remains to indicate how a theory of truth can be applied to particular speakers or groups of speakers. Given the complexity of the structures the concept of truth helps characterize, comparatively anaemic bits of evidence, applied at a potential infinity of points, can yield rich and instructive results. But complete formalization of the relation between evidence for the theory and the theory itself is not to be expected.

What we *should* demand, however, is that the evidence for the theory be in principle publicly accessible, and that it not assume in advance the concepts to be illuminated. The requirement that the evidence be publicly accessible is not due to an atavistic yearning for behavioristic or verificationist foundations, but to the fact that what is to be explained is a social phenomenon. Mental phenomena in general may or may not be private, but the correct interpretation of one person's speech by another must *in principle* be possible. A speaker's intention that her words be understood in a certain way may of course remain opaque to the most skilled and knowledgeable listener, but what has to do with correct interpretation, meaning, and truth conditions is necessarily based on available evidence. As Ludwig Wittgenstein, not to mention Dewey, G. H. Mead, Quine, and many others have insisted, language is intrinsically social. This does not entail that truth and meaning can be *defined* in terms of observable behavior, or that it is "nothing but" observable behavior; but it does imply that meaning is entirely determined by observable behavior, even readily observable behavior. That meanings are decipherable is not a matter of luck; public availability is a constitutive aspect of language.

[58] I explained in the previous section why I believe we do not need to worry separately about reference or satisfaction. Put briefly, the reason is that T-sentences do not contain a referential concept. Since the testable implications of the theory are T-sentences as applied to cases, any way of characterizing satisfaction that yields confirmable T-sentences is as good as another.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The concepts used to express the evidence must not beg the question; they should be sufficiently remote from what the theory ultimately produces. This final condition is no more than we ask of any revealing analysis, but it is difficult, at least in this case, to satisfy it. Any attempt to understand verbal communication must view it in its natural setting as part of a larger enterprise. It seems at first that this cannot be difficult, there being no more to language than public transactions among speakers and interpreters, and the aptitudes for such transactions. Yet the task eludes us. For the fact that linguistic phenomena are nothing but behavioral, biological, or physical phenomena described in an exotic vocabulary of meaning, reference, truth, assertion, and so on—mere supervenience of this sort of one kind of fact or description on another—does not guarantee, or even hold out promise of, the possibility of conceptual reduction.

There lies our problem. I shall now sketch what I think is at least the right *sort* of solution. The immediate psychological environment of linguistic aptitudes and accomplishments is to be found in the attitudes, states, and events that are described in intensional idiom: intentional action, desires, beliefs, and their close relatives like hopes, fears, wishes, and attempts. Not only do the various propositional attitudes and their conceptual attendants form the setting in which speech occurs, but there is no chance of arriving at a deep understanding of linguistic facts except as that understanding is accompanied by an interlocking account of the central cognitive and conative attitudes.

It is too much to ask that these basic intensional notions be reduced to something else—something more behavioristic, neurological, or physiological, for example. Nor can we analyze any of these basic three—belief, desire, and meaning—in terms of one or two of the others; or so I think, and have argued elsewhere.[59] But even if we could effect a reduction in this basic trio, the results would fall short of what might be wanted simply because the end point—the interpretation, say, of speech—would be too close to where we began (with belief and desire, or with intention, which is the product of belief and desire). A basic account of any of these concepts must start beyond or beneath them all, or at some point equidistant from them all.

If this is so, an analysis of linguistic meaning that assumes prior identification of nonlinguistic purposes or intentions will be radically

[59] For considerations in support of these claims, see my "Belief and the Basis of Meaning," *Synthese*, XXVII (1974): 309–323; "Radical Interpretation," *Dialectica*, XXVII (1973): 313–328; and "Thought and Talk," in Samuel Guttenplan, ed., *Mind and Language* (New York: Oxford, 1975), pp. 7–23.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

incomplete. Nor will it help to appeal to explicit or implicit rules or conventions, if only because these must be understood in terms of intentions and beliefs. Conventions and rules do not explain language; language explains them. There is no doubting, of course, the importance of showing how meanings and intentions are connected. Such connections give structure to the propositional attitudes and suit them to systematic treatment. But the interdependence of the basic intentional attitudes is so complete that it is bootless to hope to understand one independently of understanding the others. What is wanted, then, is an approach that yields an interpretation of a speaker's words at the same time that it provides a basis for attributing beliefs and desires to the speaker. Such an approach aims to provide a basis for, rather than to assume, the individuation of propositional attitudes.

Bayesian decision theory, as developed by Ramsey,[60] deals with two of the three intentional aspects of rationality that seem most fundamental, belief and desire. The choice of one course of action over another, or the preference that one state of affairs obtain rather than another, is the product of two considerations: the value placed on the various possible consequences, and how likely those consequences are deemed to be, given that the action is performed or the state of affairs comes to obtain. In choosing an action or states of affairs, therefore, a rational agent will select one, the relative value of whose possible outcomes, when tempered by the probability the agent assigns to those consequences, is greatest. Acting is always a gamble, since an agent can never be certain how things will turn out. So to the extent that an agent is rational he will take what he believes is the best bet available (he "maximizes expected utility").

A feature of such a theory is that what it is designed to explain—ordinal preferences or choices among options—is relatively open to observation, while the explanatory mechanism, which involves degree of belief and cardinal values, is not taken to be observable. The issue therefore arises of how to tell when a person has a certain degree of belief in some proposition, or what the relative strengths of his preferences are. The evident problem is that what is known (ordinal, or simple preference) is the resultant of two unknowns, degree of belief and relative strength of preference. If a person's cardinal preferences for outcomes were known, then his choices

[60] "Truth and Probability," in *The Foundations of Mathematics* (New York: Humanities, 1950), pp. 156–198.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

among courses of action would reveal his degree of belief; and if his degree of belief were known, his choices would disclose the comparative values he puts on the outcomes. But how can both unknowns be determined from simple choices or preference alone? Ramsey solved this problem by showing how, on the basis of simple choices alone, it is possible to find a proposition treated as being as likely to be true as its negation. This single proposition can then be used to construct an endless series of wagers, choices among which yield a measure of value for all possible options and eventualities. It is then routine to calculate the degrees of belief in all propositions.

Ramsey was able to turn this trick by specifying constraints on the permissible patterns of simple preferences or choices. These constraints are not arbitrary, but are part of a satisfactory account of the reasons for a person's preferences and choice behavior. The constraints spell out the demand that an agent be rational, not in his particular and ultimate values, but in the pattern these form with one another and in combination with his beliefs. The theory thus has a strong normative element, but an element that is essential if the concepts of preference, belief, reason, and intentional action are to have application.

Pattern in what is observed is central to the intelligibility of an agent's choice behavior—it determines our ability to understand actions as done for a reason. The same pattern is central to the theory's power to extract, from facts that taken singly are relatively directly connected with what can be observed, facts of a more sophisticated kind (degree of belief, comparisons of differences in value). From the point of view of the theory, the sophisticated facts explain the simple, more observable ones, while the observable ones constitute the evidential base for testing or applying the theory.

Bayesian decision theory does not provide a definition of the concepts of belief and preference on the basis of nonintensional notions. Rather, it makes use of one intensional notion, ordinal preference between gambles or outcomes, to give content to two further notions, degree of belief and comparisons of differences in value. So it would be a mistake to think the theory provides a reduction of intensional concepts to something else. Nevertheless, it is an important step in the direction of reducing complex and relatively theoretical intensional concepts to intensional concepts that in application are closer to publicly observable behavior. Above all, the theory shows how it is possible to assign a content to two basic and interlocking propositional attitudes without assuming that either one is understood in advance.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

As a theory for explaining human actions, a Bayesian decision theory of the sort I have been describing is open to the criticism that it presupposes that we can identify and individuate the propositions to which attitudes like belief and desire (or preference) are directed. But as urged a few pages back, our ability to identify, and distinguish among, the propositions an agent entertains is not to be separated from our ability to understand what he says. We generally find out exactly what someone wants, prefers, or believes only by interpreting his speech. This is particularly obvious in the case of decision theory, where the objects to be chosen or preferred must often be complex wagers, with outcomes described as contingent on the occurrence of specific events. Clearly, a theory that attempts to elicit the attitudes and beliefs that explain preferences or choices must include a theory of verbal interpretation if it is not to make crippling assumptions.

What we must add to decision theory, or incorporate in it, is a theory of verbal interpretation, a way of telling what an agent means by his words. Yet this addition must be made in the absence of detailed information about the propositional contents of beliefs, desires, or intentions.

In important respects, Quine's approach to meaning is strikingly similar to Ramsey's approach to decision making. Noting that, while there is no direct way to observe what speakers mean, all the evidence required to implement communication must be publicly available, Quine surveys the relevant available evidence, and asks how it could be used to elicit meanings. What can be observed, of course, is speech behavior in relation to the environment, and from this certain attitudes toward sentences can be fairly directly inferred, just as preferences can be inferred from choices. For Quine, the key observables are acts of *assent* and *dissent*, as caused by events within the ambit of the speaker. From such acts it is possible to infer that the speaker is caused by certain kinds of events to hold a sentence true.[61]

Just here a basic challenge arises. A speaker holds a sentence true as a result of two considerations: what he takes the sentence to mean, and what he believes to be the case. The problem is that what is relatively directly observable by an interpreter is the product of two unobservable attitudes, belief and meaning. How can the roles of these two explanatory factors be distinguished and extracted from the evidence? The problem is curiously like the problem of disentan-

[61] The step from observed assents to the inferred attitude of holding true is not, I think, explicit in Quine.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

gling the roles of belief and preference in determining choices and preferences.

Quine's solution resembles Ramsey's, in principle if not in detail. The crucial step in both cases is to find a way to hold one factor steady in certain situations while determining the other. Quine's key idea is that the correct interpretation of an agent by another cannot intelligibly admit certain kinds and degrees of difference between interpreter and interpreted with respect to belief. As a result, an interpreter is justified in making certain assumptions about the beliefs of an agent before interpretation begins. As a constraint on interpretation, this is often called by the name Neil Wilson[62] gave it, the *principle of charity*. As a device for separating meaning and belief without assuming either, it is a brilliant alternative to any approach to meaning that takes meanings for granted or assumes the analytic-synthetic distinction.

In what follows, I use Quine's inspired method in ways that deviate, sometimes substantially, from his. A difference relevant to the present topic is this. Where Quine is concerned with the conditions of successful translation from a speaker's language into the interpreter's, I emphasize what the interpreter needs to know of the semantics of the speaker's language, that is, what is conveyed by the T-sentences entailed by a theory of truth. The relation between these two projects, Quine's and mine, is obvious; given a theory of truth for a speaker's language L stated in the interpreter's language M, it is fairly straightforward to produce a manual that translates (at least roughly) from L to M.[63] But the converse is false; there are many sentences we can translate without having any idea how to incorporate them in a theory of truth. Demanding that a theory of interpretation satisfy the constraints of a theory of truth means that more structure than is needed for translation must be made manifest.

If we suppose, as the principle of charity says we unavoidably must, that the pattern of sentences to which a speaker assents reflects the semantics of the logical constants, it is possible to detect and interpret those constants. The guiding principles here, as in decision theory, derive from normative considerations. The relations between beliefs play a decisive constitutive role; an interpreter cannot accept great or obvious deviations from his own standards of rationality

[62] "Substances Without Substrata," *Review of Metaphysics*, XII (1959): 521–539.

[63] The sailing may not be completely straight; it is easy to imagine a language which contains no translation of the English word 'now' but which can give the truth conditions of English sentences containing the word 'now'.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

without destroying the foundation of intelligibility on which all interpretation rests. The possibility of understanding the speech or actions of an agent depends on the existence of a fundamentally rational pattern, a pattern that must, in general outline, be shared by all rational creatures. We have no choice, then, but to project our own logic on to the language and beliefs of another. This means it is a constraint on possible interpretations of sentences held true that they are (within reason) logically consistent with one another.

Logical consistency yields no more than the interpretation of the logical constants, however (whatever we take to be the limits of logic and the list of logical constants). Further interpretation requires further forms of agreement between speaker and interpreter. Assuming that the identification of the logical constants required for first-order quantificational structure has been accomplished, it is possible to identify as such singular terms and predicates. This raises the question how these are to be interpreted. Here progress depends on attending, not just to which sentences an agent holds true, but also to the events and objects in the world that cause him to hold the sentences true. The circumstances, observable by speaker and interpreter alike, under which an agent is caused to accept as true sentences like 'It is raining', 'That's a horse', or 'My foot hurts', provide the most obvious evidence for the interpretation of these sentences and the predicates in them. The interpreter, on noticing that the agent regularly accepts or rejects the sentence 'The coffee is ready' when the coffee is or is not ready will (however tentatively pending related results) try for a theory of truth that says that an utterance by the agent of the sentence 'The coffee is ready' is true if and only if the coffee can be observed by the agent to be ready at the time of the utterance.

The interpretation of common predicates and names depends heavily on indexical elements in speech, such as demonstratives and tense, since it is these which most directly allow predicates and singular terms to be connected to objects and events in the world. (To accommodate indexical elements, theories of truth of the sort proposed by Tarski must be augmented; the nature of these modifications has been discussed elsewhere.[64]) The method I propose for interpreting the more observational sentences and predicates is similar in some respects to the method of Quine's *Word and Object*

[64] The sort of modification required is discussed in *Inquiries into Truth and Interpretation.*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(§§7–10), but it differs in others. The most important difference concerns the objects or events that determine communicable content. For Quine this is the patterns of nerve endings which prompt assent to a sentence; an observation sentence of a speaker is "stimulus synonymous" with an observation sentence of an interpreter if speaker and interpreter would be prompted to accept or reject their respective sentences by the same patterns of proximal stimulation. Quine's idea is to capture in respectably scientific form the empiricist idea that meaning depends on evidence directly available to each speaker. My approach is by contrast externalist: I suggest that interpretation depends (in the simplest and most basic situations) on the external objects and events salient to both speaker and interpreter, the very objects and events the speaker's words are then taken by the interpreter to have as subject matter. It is the distal stimulus that matters to interpretation.[65] The significance of this point will be assessed presently.

The difficulty with what we may call the *distal theory of reference* is that it makes error hard to explain, the crucial gap between what is believed true and what is true; since the distal theory bases truth on belief, the problem is crucial. The solution depends on two closely related interpretive devices. An interpreter bent on working out a speaker's meanings notes more than what causes assents and dissents; he notes how well placed and equipped the speaker is to observe aspects of his environment, and accordingly gives more weight to some verbal responses than to others. This provides him with the rudiments of an explanation of deviant cases where the speaker calls a sheep a goat because he is mistaken about the animal rather than the word. The subtler and more important device depends on the interanimation of sentences. By this I mean the extent to which a speaker counts the truth of one sentence as supporting the truth of others. We have seen an example of how evidence of such dependencies leads to the interpretation of logical constants. But matters of evidential support can also aid in the interpretation of so-called observational terms, by helping explain error.

The interpretation of terms less directly keyed to untutored observation must also depend in large measure on conditional probabilities, which show what the agent counts as evidence for the appli-

[65] I have discussed this aspect of Quine's theory of meaning in "Meaning, Truth and Evidence," in R. Gibson, ed., *Perspectives on Quine* (New York: Blackwell, 1989). There I note that Quine also sometimes seems to subscribe to the "distal" theory, especially in *The Roots of Reference* (La Salle, IL: Open Court, 1973).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

cation of his more theoretical predicates. If we want to identify and so interpret the role of a theoretical concept or its linguistic expression, we must know how it relates to other concepts and words. These relations are in general holistic and probabilistic. We can therefore spot them only if we can detect the degree to which an agent holds a sentence true, his subjective probabilities. Simple assent and dissent are at extreme and opposite ends of a scale; we need to be able to locate attitudes intermediate in strength. Degree of belief, however, cannot be directly diagnosed by an interpreter; as we saw in discussing decision theory, degree of belief is a construction based on more elementary attitudes.

The theory of verbal interpretation and Bayesian decision theory are evidently made for each other. Decision theory must be freed from the assumption of independent access to meanings; theory of meaning calls for a theory of degree of belief in order to make serious use of relations of evidential support. But stating these mutual dependencies is not enough, for neither theory can be developed first as a basis for the other. There is no way simply to add one to the other since in order to get started each requires an element drawn from the other. What is wanted is a unified theory that yields degree of belief, desirabilities on an interval scale, and an interpretation of speech, a theory that does not assume that either desires or beliefs have been individuated in advance, much less quantified.

Such a theory must be based on some simple attitude that an interpreter can recognize in an agent before the interpreter has detailed knowledge of any of the agent's propositional attitudes. The following attitude will serve: the attitude an agent has toward two of his sentences when he prefers the truth of one to the truth of the other. The sentences must be endowed with meaning by the speaker, of course, but interpreting the sentences is part of the interpreter's task. What the interpreter has to go on, then, is information about what episodes and situations in the world cause an agent to prefer that one rather than another sentence be true. Clearly an interpreter can know this without knowing what the sentences mean, what states of affairs the agent values, or what he believes. But the preferring true of sentences by an agent is equally clearly a function of what the agent takes the sentences to mean, the value he sets on various possible or actual states of the world, and the probability he attaches to those states contingent on the truth of the relevant sentences. So it is not absurd to think that all three attitudes of the agent can be abstracted from the pattern of an agent's preferences among sentences.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

It may be objected that a preference for the truth of one sentence rather than another is itself an intentional state, and one that could be known to hold only on the assumption that many psychological factors are present. This is true (as it is also of assent to, or the holding true of, a sentence). But the objective was not to avoid intentional states; it was to avoid *individuative* intentional states, *intensional* states, states with (as one says) a propositional object. A preference for the truth of one sentence over another is an extensional relation that relates an agent and two sentences (and a time). Because it can be detected without knowing what the sentences mean, a theory of interpretation based on it can hope to make the crucial step from the nonpropositional to the propositional.

Here, in outline, is how I think the hope can be satisfied. We have already seen (again in survey form) how to arrive at a theory of meaning and belief on the basis of knowledge about the degrees to which sentences are held true. So if we could derive degree of belief in sentences by appeal to information about preferences that sentences be true, we would have a successful unified theory.

Ramsey's version of Bayesian decision theory makes essential use of gambles or wagers, and this creates a difficulty for my project. For how can we tell that an agent views a sentence as presenting a gamble until we are far along in the process of interpreting his language? A gamble, after all, specifies a connection, presumably causal, between the occurrence of a certain event (a coin comes up heads) and a specific outcome (you win a horse). Even assuming we could tell when an agent accepts such a connection, straightforward application of the theory depends also on the causing event (the coin coming up heads) having no value, negative or positive, in itself. It is also necessary to assume that the probability the agent assigns to the coin coming up heads is not contaminated by thoughts about the likelihood of winning a horse. In experimental tests of decision theory, one tries to provide environments in which these assumptions have a chance of being true; but the general application we now have in mind cannot be so choosy.

We owe to Richard Jeffrey[66] a version of Bayesian decision theory

[66] *The Logic of Decision* (Chicago: University Press, 2nd ed., 1983). Jeffrey's theory does not determine probabilities and utilities up to the same sets of transformations as standard theories. Instead of a utility function determined up to a linear transformation, in Jeffrey's theory the utility function is unique only up to a fractional linear transformation; and the probability assignments, instead of being unique once a number is chosen for measuring certainty (always One), are unique only to within a certain quantization. These diminutions in determinacy are con-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that makes no direct use of gambles, but treats the objects of preference, the objects to which subjective probabilities are assigned, and the objects to which relative values are assigned all as propositions. Jeffrey has shown in detail how to extract subjective probabilities and values from preferences that propositions be true.

An obvious problem remains. Jeffrey shows how to get results much like Ramsey's by substituting preferences among propositions for preferences among gambles. But propositions are meanings, or sentences with meanings, and if we know the propositions an agent is choosing among, our original problem of interpreting language and individuating propositional attitudes is assumed to have been solved from the start. What we want is to get Jeffrey's results, but starting with preferences among uninterpreted sentences, not propositions.

This turns out to be a soluble problem. Jeffrey's method for finding the subjective probabilities and relative desirabilities of propositions depends only on the truth-functional structure of propositions —on how propositions are made up out of simple propositions by repeated application of conjunction, disjunction, negation, and the other operations definable in terms of these. If we start with sentences instead of propositions, then the crucial difficulty will be overcome provided the truth-functional connectives can be identified. For once the truth-functional connectives have been identified, Jeffrey has shown how to fix, to the desired degree, the subjective desirabilities and probabilities of all sentences; and this, I have argued, suffices to yield a theory for interpreting the sentences. Knowing an agent's evaluative and cognitive attitudes to interpreted sentences is not (at least in the context of this approach) to be distinguished from knowing the agent's beliefs and desires. The essential steps in this procedure, particularly the procedure that elicits the interpretation of the truth-functional connectives from facts about preference, are described in the appendix to this essay.

The approach to the problems of meaning, belief, and desire which I have outlined is not, I am sure it is clear, meant to throw any direct light on how in real life we come to understand each other, nor

---

ceptually and practically appropriate: they amount, among other things, to permitting somewhat the same sort of indeterminacy in decision theory that we have come to expect in a theory of linguistic interpretation. Just as you can account for the same data in decision theory using various utility functions by making corresponding changes in the probability function, so you can change the meanings you attribute to a person's words (within limits) provided you make compensating changes in the beliefs you attribute to him.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

on how we master our first concepts and our first language.[67] I have been engaged in a conceptual exercise aimed at revealing the dependencies among our basic propositional attitudes at a level fundamental enough to avoid the assumption that we can come to grasp them—or intelligibly attribute them to others—one at a time. Performing the exercise has required showing how it is in principle possible to arrive at all of them at once. Showing this amounts to presenting an informal proof that we have endowed thought, desire, and speech with a structure that makes interpretation possible. Of course, we knew it was possible in advance. The philosophical question was, *what* makes it possible?

What makes the task practicable at all is the structure the normative character of thought, desire, speech, and action imposes on correct attributions of attitudes to others, and hence on interpretations of their speech and explanations of their actions. What I have said about the norms that govern our theories of intensional attribution is crude, vague, and incomplete. The way to improve our understanding of such understanding is to improve our grasp of the standards of rationality implicit in all interpretation of thought and action.

The idea that the propositional content of observation sentences is (in most cases) determined by what is common and salient to both speaker and interpreter is a direct correlate of the common-sense view of language learning. It has profound consequences for the relation between thought and meaning, and for our view of the role of truth, for it not only ensures that there is a ground level on which speakers share views, but also that what they share is a largely correct picture of a common world. The ultimate source of both objectivity and communication is the triangle that, by relating speaker, interpreter, and the world, determines the contents of thought and speech. Given this source, there is no room for a relativized concept of truth.

[67] Given the intricacy of any interpretable system of thought and language, I assume that there must be many alternative approaches to interpretation. I have outlined one; others may well be less artificial or closer to our intuitions concerning interpretive practice. But one should not take for granted that the procedure I have sketched is totally remote from what is practicable. As a start, observe that every utterance that can be treated as a sincere request or demand may be taken to express the utterer's preference that a certain sentence be true rather than its negation. Most experimental work in decision theory takes as data the choices subjects make between alternatives as described in writing or speech. It is normally assumed that subjects understand these descriptions as the experimenters do. Dropping this assumption yields data of exactly the sort required by the approach presented here.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

We recognized that truth must somehow be related to the attitudes of rational creatures; this relation is now revealed as springing from the nature of interpersonal understanding. Linguistic communication, the indispensable instrument of fine-grained interpersonal understanding, rests on mutually understood utterances, the contents of which are finally fixed by the patterns and causes of sentences held true. The conceptual underpinning of interpretation is a theory of truth; truth thus rests in the end on belief and, even more ultimately, on the affective attitudes.

DONALD DAVIDSON

University of California/Berkeley

APPENDIX

Jeffrey's method for finding the subjective probabilities and relative desirabilities of propositions depends only on the truth-functional structure of propositions—on how propositions are made up out of simple propositions by repeated application of conjunction, disjunction, negation, and the other operations definable in terms of these. If we start with sentences instead of propositions, then our problem will be solved provided the truth-functional connectives can be identified. For once the truth-functional connectives have been identified, Jeffrey has shown how to fix, to the desired degree, the subjective desirabilities and probabilities of all sentences; and this, I have argued, suffices to yield a theory for interpreting the sentences.

The basic empirical primitive in the method to be described is the agent's (weak) preference that one sentence rather than another be true; one may therefore think of the data as being of the same sort as the data usually gathered in an experimental test of any Bayesian theory of decision, with the proviso that the interpretation of the sentences among which the agent chooses is not assumed known in advance to the interpreter.

The uniformity and simplicity of the empirical ontology of the system, comprising as it does just utterances and sentences, is essential to achieving the aim of combining decision theory with interpretation. I shall follow Jeffrey, whose theory deals with propositions only, as closely as possible, substituting uninterpreted sentences where he assumes propositions. Here, then, is the analogue of Jeffrey's *desirability axiom* (D), applied to sentences rather than propositions:

(D) If $prob(s \text{ and } t) = 0$ and $prob(s \text{ or } t) \neq 0$, then

$$des(s \text{ or } t) = \frac{prob(s)des(s) + prob(t)des(t)}{prob(s) + prob(t)}$$

(I write '$prob(s)$' for the subjective probability of s and '$des(s)$' for the desirability of s.) By relating preference and belief, this axiom does the sort of work usually done by gambles; the relation is, however, different. Events are correlated with sentences that on interpretation turn out to say the event occurs ('The next card is a club'). Actions and outcomes are also represented by sentences ('The agent bets one dollar', 'The agent wins five dollars'). Gambles do not enter directly, but the element of risk is present, since to choose that a sentence be true is usually to take a risk on what will concomitantly be true. (It is assumed that one cannot choose a logically false sentence.) So we see that, if the agent chooses to make true rather than false the sentence 'The agent bets one dollar', he is taking a chance on the outcome, which may, for example, be thought to depend on whether or not the next card is a club. Then the desirability of the (truth of) the sentence 'The agent bets one dollar' will be the desirability of the various circumstances in which this sentence is true,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

weighted in the usual way by the probabilities of those circumstances. Suppose the agent believes he will win five dollars if the next card is a club and will win nothing if the next card is not a club; he will then have a special interest in whether the truth of 'The agent bets one dollar' will be paired with the truth or falsity of 'The next card is a club'. Let these two sentences by abbreviated by 's' and 't'. Then

$$des(s) = \frac{prob(s \text{ and } t)des(s \text{ and } t) + prob(s \text{ and } \sim t)des(s \text{ and } \sim t)}{prob(s)}$$

This is, of course, something like Ramsey's gambles. It differs, however, in that there is no assumption that the "states of nature" that may be thought to determine outcomes are, in Ramsey's term, "morally neutral," that is, have no effect on the desirabilities of the outcomes. Nor is there the assumption that the probabilities of outcomes depend on nothing but the probabilities of the "states of nature" (the agent may believe he has a chance of winning five dollars even if the next card is not a club, and a chance he will not win five dollars even if the next card is a club).

The desirability axiom can be used to show how probabilities depend on desirabilities in Jeffrey's system. Take the special case where $t = \sim s$. Then we have

(1) $$des(s \text{ or } \sim s) = des(s)prob(s) + des(\sim s)prob(\sim s)$$

Since $prob(s) + prob(\sim s) = 1$, we can solve for $prob(s)$:

(2) $$prob(s) = \frac{des(s \text{ or } \sim s) - des(\sim s)}{des(s) - des(\sim s)}$$

Thus, the probability of a proposition depends on the desirability of that proposition and of its negation. Further, it is easy to see that, if a sentence s is more desirable than an arbitrary logical truth (such as 't or $\sim t$'), then its negation ('$\sim s$') cannot also be more desirable than a logical truth. Suppose that we assign the number 0 to any logical truth. (This is intuitively reasonable since an agent is indifferent to the truth of a tautology.) Then (2) can be rewritten:

(3) $$prob(s) = \frac{1}{1 - \dfrac{des(s)}{des(\sim s)}}$$

It is at once apparent that $des(s)$ and $des(\sim s)$ cannot both be more, or both be less, than 0, the desirability of any logical truth, if $prob(s)$ is to fall in the interval from 0 to 1. If (following Jeffrey) we call an option good if it is preferred to a logical truth and bad if a logical truth is preferred to it, then (3) shows that it is impossible for an option (sentence) and its negation both to be good or both to be bad.

Taking '$\sim(s \text{ and } \sim s)$' as our sample logical truth, we can state this principle purely in terms of preferences:

(4) If $des(s) > des(\sim(s \text{ and } \sim s))$ then

$$des(\sim(s \text{ and } \sim s)) \geq des(\sim s), \text{ and}$$

if $des(\sim(s \text{ and } \sim s)) > des(s)$ then

$$des(\sim s) \geq des(\sim(s \text{ and } \sim s)).$$

Since both negation and conjunction can be defined in terms of the Sheffer stroke '|' ("not both"), (4) can be rewritten:

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(5) If $des(s) > des((t|u)|((t|u)|(t|u)))$ then

$$des((t|u)|((t|u)|(t|u))) \geq des(s|s), \text{ and}$$

if $des((t|u)|((t|u)|(t|u))) > des(s)$ then

$$des(s|s) \geq ((des((t|u)|((t|u)|(t|u))).$$

The interest of (5) for present purposes is this. If we assume that '|' is some arbitrary truth-functional operator that forms sentences out of pairs of sentences, then the following holds: if (5) is true for all sentences s, t, and u, and for some s and t, $des(s|s) \neq des(t|t)$, then '|' must be the Sheffer stroke (it must have the logical properties of "not both"); no other interpretation is possible.[68]

Thus, data involving only preferences among sentences, the meanings of which are unknown to the interpreter, have led (given the constraints of the theory) to the identification of one sentential connective. Since all logically equivalent sentences are equal in desirability, it is now possible to interpret all the other truth-functional sentential connectives, since all are definable in terms of the Sheffer stroke. For example, if it is found that, for all sentences s,

$$des(s|s) = des(\sim s)$$

we can conclude that the tilde is the sign for negation.

It is now possible to measure the desirability and subjective probability of all sentences, for the application of formulas like (2) and (3) requires the identification of only the truth-functional sentential connectives. Thus, it is clear from (3) that, if two sentences are equal in desirability (and are preferred to a logical truth) and their negations are also equal in desirability, the sentences must have the same probability. By the same token, if two sentences are equal in desirability (and are preferred to a logical truth), but the negation of one is preferred to the negation of the other, then the probability of the first is less than that of the second. This, along with appropriate existence axioms, is enough to establish a probability scale. Then it is easy to determine the relative desirabilities of all sentences.[69]

At this point the probabilities and desirabilities of all sentences have in theory been determined. But no complete sentence has yet been interpreted, though the truth-functional sentential connectives have been identified, and so sentences logically true or false by virtue of sentential logic can be recognized.

We have shown how to interpret the simplest sentences on the basis of (degrees of) belief in their truth, however. Given degrees of belief and the relative strengths of desire for the truth of interpreted sentences, we can give a propositional content to the beliefs and desires of an agent.

[68] I am indebted to Stig Kanger for showing me why an earlier attempt at a solution to this problem would not work. He also added some needed refinements to the present proposal.

[69] For details see Jeffrey, *The Logic of Decision*.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

TODAY'S ETHICAL ISSUES

THE HUMAN RIGHTS READER

Revised Edition

Edited by Walter Laqueur and Barry Rubin. This unique documentary history traces the philosophical, political and legal evolution of the concept of human rights. This expanded edition adds key speeches and policy statements by individuals and organizations including Andrei Sakharov, Bishop Desmond Tutu and Amnesty International and includes an extensive bibliography.

MERIDIAN
ISBN: 0-452-01026-8 $10.95
January 1990

Prices subject to change. Write to the NAL Education Department at the address below for a free Psychology and Social Sciences catalog.

THE CRISIS IN HEALTH CARE

Ethical Issues

Edited by Nancy F. McKenzie, Ph.D. Provocative essays by such authorities as Alan Brandt, Barbara and John Ehrenreich and Hans Jonas address the most significant social, political and medical issues involving American health care.

MERIDIAN
ISBN: 0-452-01028-4 $12.95 June 1990

ESSENTIALS OF BUSINESS ETHICS

Edited by Peter Madsen, Ph.D. and Jay Shafritz, Ph.D. A comprehensive look at the evolution of ethics and its role in today's business world. The collection covers such controversial topics as drug testing and environmental responsibility and includes articles by Ralph Nader, Milton Friedman and Peter Drucker.

MERIDIAN
ISBN: 0-452-01044-6 $9.95 May 1990

NAL
NEW AMERICAN LIBRARY
A Division of Penguin, USA
1633 Broadway, New York, NY 10019

the philosophical forum
A QUARTERLY

VOLUME XXI:1-2
Fall-Winter 1989-90 $8.50

HERMENEUTICS IN ETHICS AND SOCIAL THEORY
A SPECIAL DOUBLE ISSUE

Jürgen Habermas
Michael Walzer
Agnes Heller
Thomas McCarthy
Seyla Benhabib
Albrecht Wellmer
and others

Editor: Marx Wartofsky

Guest Co-Editor:
Michael Kelly

SUBSCRIPTIONS
$15 per year for individuals
$45 per year for institutions
+ $4 for foreign subscriptions

Other special issues:
Latin American Philosophy
Fichte
Apartheid

the philosophical forum
P.O. Box 239
Baruch College of CUNY
17 Lexington Avenue
New York, NY 10010

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

MEMO TO AUTHORS

The editors cannot afford to send to the printer manuscripts that are difficult for the compositor to read or footnotes that need to be completed in proof. They ask that manuscripts submitted be prepared according to the following guidelines:

1. All material, including the quotations and footnotes—especially footnotes—must be typed, double-spaced, on one side of the paper, with ample margins left and right, top and bottom.
2. No bibliographies will be published; bibliographical material should be put into the footnotes.
3. Footnotes should be complete, according to *Journal* style (see any issue). Journal citations should include name of journal in full, volume number, issue number, month or season, year, and pages. Book references should include publisher, city of publication, and date. Footnotes should be gathered at the end.
4. Quotations should be checked by the author; capitalization, italics, spelling, etc., should match the original. Corrections should not have to be made (expensively) in proof.
5. If there are diagrams, they should be provided camera-ready by the author. If there are special symbols, they should be widely spaced and clearly marked; if there are many, a list should be provided. Standard logic symbols are available, and letters of the Greek and script alphabets; no boldface.
6. Computer copy can be extremely difficult to read; only letter-quality copy is acceptable. It should be double-spaced and not justified. The print should be dark enough and the letters widely enough spaced so that both letters and punctuation marks can be read easily.
7. Finally, photo copies on shiny paper are not acceptable. Ribbon copy is much preferred, but clear copies on bond paper are acceptable.

The editors are willing to read and evaluate manuscripts that do not meet these standards, provided they are genuinely legible and not unreasonably long. They ask for revision of excessively long manuscripts, and they require retyping of accepted manuscripts, if necessary, before scheduling for publication.

The *Journal of Philosophy* is published monthly, in issues that (ordinarily) contain only 56 pages of copy; this limits the length of papers the editors can normally consider to no more than 7,500 words, or eighteen JOURNAL pages. A more precise way of determining length is by characters: 2,850 characters per Journal page, or 51,300 characters per Journal article of 18 pages. [Calculate character count by counting the number of characters per line (including blank spaces and punctuation), multiplying by the number of lines on an average page, multiplying by the number of pages (footnotes included), and dividing by 2850.] Manuscripts very much in excess of these limits will be returned unread.

The *Journal of Philosophy* does not publish papers that are primarily historical or expository; the editors urge that such manuscripts be submitted elsewhere.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVII, NO. 7, JULY 1990

UTILITARIANISM, SOCIOBIOLOGY, AND THE LIMITS OF BENEVOLENCE

A familiar objection to act utilitarianism is that the theory ought to be rejected because it cannot be reconciled with our considered moral judgments about the extent of our duty to help others. Act utilitarianism says that we ought to maximize group utility even if that means great self-sacrifice, but quite often our belief is that such sacrifice is supererogatory, not obligatory. Another objection to act utilitarianism is that the theory ought to be rejected because the vast majority of people lack the motivations that anyone trying to live up to the theory's prescriptions would need. Even the conscientious utilitarian, it is claimed, will lack the motivations necessary to avoid frequent backsliding. The two types of objection are distinguishable only on the assumption that the utilitarian's considered judgments about what she has a duty to do cannot by themselves cause the behavior necessary to live up to them. The assumption, to put it metaphorically, seems to be that value judgments only provide a "target" and that the motivational "energy" necessary to hit the target must come from elsewhere. The first objection alleges that most of us do not aim at the utilitarian target, and the second alleges that even those who do lack the "energy" necessary to hit it consistently. John Rawls's[1] claim that the rational contractors in his original position would reject utilitarianism because they realize that it would impose excessive "strains of commitment" on them seems to be an example of the second type of objec-

[1] *A Theory of Justice* (Cambridge: Harvard, 1971), pp. 176–9. Rawls says that the utilitarian "conception of justice is threatened with instability unless sympathy and benevolence can be widely and intensely cultivated" (p. 178). He appears to believe that the amount of benevolence it requires of us "exceeds the capacity of human nature" (p. 176).

0022-362X/90/8707/329–345

© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tion. More recently, J. L. Mackie[2] and Howard Kahane[3] have argued that act utilitarianism is an "unworkable" moral theory because it demands too much benevolence of us. They claim to find support for their objection in sociobiological speculations about the basis of egoism, benevolence, and reciprocity. In the first part of this paper, I reconstruct the argument on which this second type of objection rests. My aim is to identify an argument which rests on a plausible moral psychology and which makes sociobiology at least relevant to the explanation and justification of moral beliefs. I shall show that the argument fails even if strong benevolence is beyond human nature and sociobiology correctly explains why. It fails because it confuses the concepts of "utility maximizer" and "utilitarian saint." But, as we shall see, the argument can be recast as an objection to the (nonutilitarian) principle that one ought to be a utilitarian saint, the pacifist claim that the use of force is never permissible (even in self-defense), and the Christian injunction to turn the other cheek. This recast argument (as well as the original argument against utilitarianism proper) requires "psychological determinism" and the "ought implies can" principle as two of its premises. In the latter part of this paper, I examine the version of psychological determinism needed by these arguments and presupposed by any attempt to provide a sociobiological explanation of existing moral beliefs. The recast argument does not suffer from any clear flaws (as the original one does), and my conclusion is that this argument may well be successful if the "ought implies can" principle is true.

THE ARGUMENT AGAINST UTILITARIANISM AND WHY IT FAILS

On the face of it, it is not obvious why the failure of people to fulfill the requirements of a moral principle should be regarded as a mark against the principle. Some moral theories, like Kant's, assume that certain motivations have intrinsic moral value. A genuinely moral person will act out of respect for the moral law and not from fear of God, a desire for personal gain, or "philanthropic" motives. Now, if people do not in fact act from the motivation that Kant favors, that is surely no objection to his theory. Kant could say that, if everyone acts in accordance with the moral law for egoistic reasons (like the grocer in his *Groundwork*), that only shows that everyone's actions lack moral worth, not that his moral theory is defective. The same reply

[2] "The Law of the Jungle: Moral Alternatives and the Principles of Evolution" and "Cooperation, Competition, and Moral Philosophy," in his *Persons and Values, Vol. II* (New York: Oxford, 1985), pp. 120–131 and 152–169, especially pp. 164–9.

[3] "Making the World Safe for Reciprocity," in Joel Feinberg, ed., *Reason and Responsibility* (Belmont, CA: Wadsworth, 1985, 6th ed.), pp. 495–502.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

can be made in behalf of theories that focus on behavior or states of affairs and assume that any motivation that will consistently produce the required behavior or valuable states of affairs is as good as any other motivation with that effect. According to act utilitarianism, strongly benevolent people will produce more good than people with other motivations will. If everyone were so deficient in benevolence that they seldom produced much happiness for others, that is no embarrassment to act utilitarianism. The utilitarian can say that that only shows how immoral everyone is.

This reply makes the familiar point that there is a gap between "is" and "ought." Many sociobiologists who have discussed ethical issues have been accused of ignoring this gap and committing the naturalistic fallacy.[4] This is a fair criticism of many of E. O. Wilson's pronouncements about ethics and biology. But a gap between "is" and "ought" hardly entails that the empirical claims that sociobiology may be able to support have no bearing on how we should act or which moral principles we should accept. A response to the reply of the previous paragraph is that, if someone does not have and cannot acquire the motivations required by a moral theory (or causally necessary to perform the actions required by the theory), then, granted the principle of "ought implies can," it is false that the person ought to do what the theory enjoins. Sociobiology might play a role in this response by explaining why the motivations necessary to fulfill the "oughts" of the theory are beyond human nature. The second objection to act utilitarianism can be viewed as resting on the following argument:

(a) Act utilitarianism says that one ought to be a utility maximizer—someone who performs the act that produces a larger net balance of happiness over unhappiness than is produced by any other act that is available to her.

(b) One has a moral duty to perform some act or be the sort of person disposed to perform certain acts only if one can do so ("ought implies can").

(c) If one has a duty to maximize utility, then there will be times when one has a duty to make large sacrifices (time, wealth, even one's life) for the benefit of complete strangers.

(d) One can perform any action whatsoever only if one has motivational energy of the appropriate kind and strength; possession of the motiva-

[4] Cf. Peter Singer, *The Expanding Circle* (New York: Harper & Row, 1981), pp. 72–86; and Owen J. Flanagan, "Is Morality Epiphenomenal? The Sociobiological Reduction of Ethics," *The Philosophical Forum*, XIII, 2–3 (Winter-Spring 1981–82): 207–225.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tional energy is causally necessary for performance of the action ("psychological determinism").

(e) Only extremely strong, wide-ranging benevolence could provide one with the motivational energy necessary to make the large personal sacrifices referred to in (c).

(f) Such benevolence is rare among normal human beings (ones whose personality traits have not been shaped by a eugenics program or genetic engineering), and sociobiology correctly explains why this is so.

(g) There is nothing one can do significantly to increase the strength of one's benevolence energies.

(h) Therefore, it is not true (at least of the vast majority of people) that one ought to be a utility maximizer.

It is not clear that this is the argument that underlies Mackie's charge that act utilitarianism is an "unworkable" morality, and it is highly doubtful that Rawls's "strains of commitment" objection to utilitarianism rests on it. Both simply state their objections without supporting argument. But even if Mackie and Rawls do have in mind a different argument, it is still worth inquiring whether this argument succeeds in refuting act utilitarianism. Since the argument is clearly valid, the utilitarian can reject the conclusion only if she rejects at least one of the premises. Which premise is most vulnerable to criticism?

The utilitarian could reject (b), the "ought implies can" principle. I shall return to the principle at the very end of this paper. For now I assume that the act utilitarian accepts it and interprets it the way the principle must be interpreted (with the "can" having a categorical rather than hypothetical sense) if the argument is to go through. In any case, she *can* accept it without admitting defeat. There are other, more obvious problems with the argument on which she should focus.

The act utilitarian might reject (c), the claim that her theory demands great self-sacrifice, for the following reasons. Edmund Burke thought (and many twentieth-century conservatives continue to think) that busybody do-gooders bent on social reform tend to do more harm than do people who mind their own business. Thomas Malthus thought that charity for the poor, weak, and hungry does more harm than good in the long run because it only reinforces laziness and dependence on others, and more importantly, because it only produces more hungry mouths in the future. At the heart of such conservative thinking lies the faith that there are "invisible hands" at work in society—hands which guarantee that millions of uncoordinated egoistic acts will have better consequences than mil-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

lions of semi-coordinated self-sacrificing acts would. Although such faith was common among utilitarians in the eighteenth and nineteenth centuries, it is not so common today. Once it is recognized that there are at least as many prisoner's dilemmas in the social world as there are invisible hands, it becomes hard to sustain the faith that this objection to (c) requires.

A more plausible objection (with some similarities to the one just dismissed) is provided by R. M. Hare's[5] "two-level" version of act utilitarianism. In a two-level theory, the principle that is the standard of rightness differs from the principle (or set of principles) used in moral deliberation to guide decision making. Hare thinks that "critical level" moral reflection, which one engages in when one is calm, unhurried, and thinking clearly, tells one that act utilitarianism is the correct standard of rightness. But it does not tell one to use this principle as one's decision procedure (which, in Hare's terminology, is to be a utilitarian at the "intuitive level"). It does not tell one to *try* to maximize utility, to sacrifice oneself or others whenever one *thinks* that the gains to the community would outweigh the losses to a few. For "critical level" moral reflection realizes that, for a number of different reasons, people who always try to maximize utility produce less total utility than people who try to be good Kantians or scrupulous followers of the Ten Commandments.

Call someone a "utilitarian saint" if, having attached as much weight to the sufferings of total strangers as she does to her own or those of friends and relatives, she follows Peter Singer's[6] advice and gives her money to those worse-off than herself until the utility for her of the last dollar she gives away equals the utility for them of the last dollar they receive. Distinguishing between the "two levels" shows us why act utilitarianism does not tell us to try to maximize utility, and hence why it certainly does not tell us to try to be utilitarian saints. In fact, Mackie concedes that the objection of unworkableness is not telling against "two-level" versions of utilitarianism.[7] He seems to think that it undercuts only "single level" theories like the ones defended by J. J. C. Smart and Singer. But not even that claim is correct.

[5] Cf. *Moral Thinking: Its Method, Levels, and Point* (New York: Oxford, 1981). For more on the distinction between a standard of rightness and a decision procedure, cf. Peter Railton, "Alienation, Consequentialism, and the Demands of Morality," *Philosophy and Public Affairs,* XIII (Spring 1984): 134–171; also, David O. Brink, "Utilitarian Morality and the Personal Point of View," this JOURNAL, LXXXIII, 8 (August 1986): 417–438.

[6] "Famine, Affluence, and Morality," *Philosophy and Public Affairs,* I, 3 (Spring 1972): 229–243.

[7] "Cooperation, Competition, and Moral Philosophy," p. 165.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

A "single-level" act utilitarianism says that often utility maximization should be one's conscious goal and that moral deliberation should consist in trying to ascertain the best means to this goal. But even this version of utilitarianism only requires one to choose the course of action that produces the most utility from those which are *available* to her. If psychological determinism is true, one's benevolence energies weak, and there is nothing one can do to increase them, that shrinks the number of actions that are available to one. Suppose that some people possess the magical power to cure diseases and stop earthquakes by snapping their fingers. They can produce a lot more utility than we can. Although act utilitarianism tells them to snap their fingers, it does not tell us to, nor does it tell us that we should try to acquire their magical powers if that would be a waste of time. Now the same point holds for saints and us. They can produce a lot more utility than we can. But if we possess limited sympathies and there is nothing we can do to change that fact about ourselves, then it would be a waste of time for us to try to do the super-benevolent things that they do or to try to acquire their motivations. (Indeed, not just time might be wasted. If we thought that we ought to make saint-like sacrifices but then did not make them, that would lead to guilt feelings. If the guilt were chronic and strong, we might become alienated from morality altogether—much as unenforceable prohibition laws may have encouraged contempt for other laws.) Instead, the act utilitarian can tell us, we should perform that act from the set of acts that people with limited sympathies can perform which will produce the most utility.

We can now summarize the problem with the argument against utilitarianism as follows. The conscientious utilitarian with strongly benevolent motivations will (other things being equal) produce more utility than the conscientious utilitarian with limited sympathies. But if the latter is producing as much utility as people with limited sympathies can, then she is as much a "utility maximizer" as the saint is. Hence, if the argument is supposed to establish the falsity of "one ought to be a utility maximizer," then it fails, because (c) is false. Utility maximizers only have the duty to make the kinds of sacrifices that people with their level of benevolence can make; utility maximizers with weak benevolence do not have the duty to do the impossible, to make the large sacrifices that (the utilitarian can concede) only saints can make. On the other hand, if "ought implies can" is granted, we lack strong benevolence and cannot get it, and psychological determinism is true, that *seems* to establish the falsity of "we ought to be utilitarian saints." The problem is that that is no objection to act utilitarianism. The reason why some people have thought

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that it is an objection is probably that they have confused the concepts of "utility maximizer" and "utilitarian saint."

BENEVOLENCE AND GENES

Utilitarianism (at least the version of it considered here) "subjectivizes" duty. It does not say "you have a duty to do *x*," but rather, "you have a duty to *try* to do *x*" or "you have a duty to bring about *as much y* as you possibly can." Many moral theories and principles do not subjectivize duty in this way. The Golden Rule does not say that I should try to treat others as I would wish they treat me, but rather that I should actually so treat them. The principle of absolute pacifism does not say that I should try never to use violence against others (even in self-defense), but that I should never use violence, period. Christianity enjoins that I love my enemies and turn the other cheek when they injure me—not just that I be as compassionate as people like me can be. The same is true of the (nonutilitarian) principle that requires that I make all the sacrifices of a utilitarian saint. The argument that we have seen fail against act utilitarianism can be restated as an argument against each of these principles. [In the new argument, (c′) says "if one has a duty to follow the Golden Rule or the Christian injunctions, to be an absolute pacifist or a utilitarian saint, then there will be times when one has a duty to make large sacrifices." (d), (e), (f), and (g) remain the same.] This new argument is no less valid than the original one. Since the moral principles attacked by this argument do not subjectivize duty, their defenders cannot sidestep the argument in the way that defenders of principles that subjectivize duty can; they have to grant the truth of (c′). This means that they must reject (b), (d), (e), (f), or (g). Which of these can they most plausibly reject?

Let us begin with an analysis of what benevolence (or altruism) is and then take a brief look at premise (f). In human beings, a first-order desire for a state of affairs will be a complex disposition which manifests itself, typically, in three ways: as a judgment in which one *approves* of the state of affairs, as a *disposition* to act in ways that one believes will bring about the state of affairs, and as various *felt inclinations* for it (e.g., savoring the prospect of its obtaining). Not all three components are equally important, and not all have to be manifested for a desire to be said to exist. Benevolence seems to be a desire in which the behavioral component is the most important part. We think that someone is benevolent if she is disposed to help another even in situations when she would not gain anything herself by helping and even when she does not believe that helping is morally required of her. The felt-inclination component of benevolence can consist in love of, affection for, or simply caring about another. If

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

someone is distressed when she finds that another is suffering, or pleased by the knowledge that another is happy, then she has some benevolence. It seems likely that such a person will be disposed to help another (at least if the costs to herself are not too high), and likely too that there is a causal connection between the disposition to help and the tendency to have these affective responses to another's joy or sorrow. Benevolence seems to be that psychological state which causes (underlies) both the disposition to help and these affective responses. It does not require that one have as a *conscious goal* increasing another's happiness or that one make the *judgment* that the increased happiness of another is valuable for its own sake. Jonathan Harrison[8] tells us that some members of Hitler's SS felt compassion for the Jews whom they executed but regarded it as their duty not to let the compassion interfere with their work. It seems correct to say that they had a trace of benevolence, even though they clearly did not disapprove of the suffering of Jews (or judge it to be morally or nonmorally bad). If, contrary to what I have claimed, benevolence requires that one intrinsically *value* the increased happiness of another, *prefer* that they be happy rather than unhappy, and *approve* of one's helping rather than ignore them, then it is doubtful that any nonhuman animals can be benevolent, because it is doubtful that any nonhuman animals have the cognitive wherewithal necessary to represent different states of affairs and *rank* them. It is only if benevolence does not require this that a bear's caring for her cubs could be due to benevolence-like motivations.

The strength of a person's benevolence is a function of several different variables. One is the size of the sacrifice that one is willing to make and the intensity of one's affective responses: other things being equal, a willingness to make a larger sacrifice and a tendency to be greatly distressed by another's suffering makes one more benevolent. Another variable is whether one needs to see the sufferer with one's own eyes to be motivated to help, or a third-hand account of her plight will do.[9] A third variable is the number one is motivated to help: other things being equal, someone motivated to help only friends and relatives, only members of her own ethnic group, or only members of her own species, is less benevolent than someone willing to help members of a larger class.

The evolutionary explanation of why strong benevolence energies are so rare among us and other mammals has been stated by the

[8] "The Conscience of Huckleberry Finn," *Philosophy*, XLIX (1974): 123–134.

[9] This point is made by Richard Brandt in "The Psychology of Benevolence and Its Implications for Philosophy," this JOURNAL, LXXIII, 14 (August 12, 1976): 429–453. Brandt also emphasizes the connection that benevolence has to both helping behavior and the affective responses to another's suffering.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

biologists William Hamilton, Robert Trivers, and John Maynard Smith. Darwinians had long been troubled by the existence of such "altruistic behavior" as the following (described by Wilson[10]):

> The dominant males of chacma baboon troops (*Papio ursinus*) position themselves in exposed locations in order to scan the environment while the other members forage. If predators of rival troops approach, the dominant males warn the others by barking and may move toward the intruders in a threatening manner, perhaps accompanied by other males. As the troop retreats, the dominant males cover the rear (*ibid.*, pp. 121–2).

A gene for such behavior surely diminishes its possessor's chances of living a long life and producing many offspring (by making it more likely that it will suffer serious injury in defense of the troop), so how does such a gene manage to get copies of itself passed on to future generations? The key to answering this question is that a creature's fitness is not just a function of the expected number of offspring that its physical traits and behavioral dispositions enable *it* to leave behind. If sacrifices for its close relatives diminish its own fecundity but sufficiently increase the fecundity of its relatives, then the sacrifices maximize its inclusive fitness and natural selection can be expected to produce creatures genetically disposed to make them. This explains "kin altruism."

It is also a mistake to suppose that natural selection could only produce creatures genetically disposed to be ruthlessly egoistic in their dealings with nonrelatives; evolution can produce "reciprocal altruism." Someone with reciprocity motivations will be helpful and feel grateful to people who have helped her, even if taking advantage of their kindness ("playing them for suckers") is in her self-interest. This differentiates reciprocity from pure egoism.[11] On the other hand, she will try to harm (or at least refuse to help) and feel vindictiveness toward those who have harmed her (or failed to reciprocate her help), even if helping (or not harming) would maximize total happiness. This differentiates reciprocity from pure benevolence. (Yet like strong benevolence, reciprocity might well lead to very large sacrifices of personal welfare: it would lead a soldier whose life has repeatedly been saved by the bravery of his comrades to risk his own life for theirs.) The biological explanation why we have reciprocity

[10] *Sociobiology: The New Synthesis* (Cambridge: Harvard, 1975) [quoted by Philip Kitcher, *Vaulting Ambition: Sociobiology and the Quest for Human Nature* (Cambridge: MIT, 1985), p. 159].

[11] One should not confuse tit-for-tat, which is a *strategy* that the rational egoist will adopt in indefinitely iterated prisoner's dilemmas, with reciprocity, which is a *basic motivation*. In iterated prisoner's dilemmas, rational egoists and reciprocators will cooperate with each other, but the motivation that produces their cooperation is very different.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

rather than strong benevolence motivations is that the former but not the latter is an "evolutionarily stable strategy." A community of genetically programmed saints could be invaded by mutant or immigrant egoists, who then flourish by taking advantage of the saints. A reciprocator, on the other hand, having been played for a sucker once by an exploitative egoist, will not allow himself to be exploited by that egoist again. Once reciprocators (with the ability to distinguish reciprocators from egoists) become numerous enough, the gene for reciprocity cannot be displaced by a gene for pure egoism. Indeed, once there are enough reciprocators in a group, it is pure, undisguised egoism that is doomed to extinction.

The sociobiologist's claim is not that strong benevolence will exist in only a few members of a group of humans, *regardless* of the environmental influences to which the whole group is or has been subject. That is too strong. Height is something commonly thought to be largely determined by genes, yet even this, everyone admits, is influenced by nutrition. The claim is presumably that, in the range of social environments in which people are likely to grow up or which could *feasibly be arranged,* strong benevolence will be rare. The sociobiologist probably wants to claim something like "unless they were sternly punished as young children when they exhibited strong egoism and rewarded with effusive praise and candy when they exhibited strong benevolence, most members of a group will have weak benevolence."

Notice that a biological explanation of why we are creatures with weak benevolence energies does not preclude the possibility of our being able to increase their strength, either directly (just by willing it) or indirectly (through self-hypnosis, behavioral modification, chemical injections, or whatever). In fact, it seems very unlikely that there is anything I could do which would transform me into someone who deeply cares about everyone else. [So premise (g) of the argument should be granted.] But whether or not this is so is a separate question from whether or not my weak benevolence has a genetic basis. Premise (g) of the argument does not follow from premise (f). We can see that 'genetically based' does not mean "inevitable" or "unalterable" if we consider short-sightedness, which sometimes has a genetic basis but which often can be corrected by wearing eyeglasses.[12] Conversely, if, as Marx thought, it is socialization into the historically transitory institutions of class-divided societies with exploitation which produces strong egoism and weak benevolence, it would hardly follow that we can significantly increase our benevo-

[12] Cf. Stephen J. Gould, *The Mismeasure of Man* (New York: Norton, 1981), p. 156.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

lence simply by establishing socialism. It might be the case that the malign effects of capitalism on our personalities are irreversible and that only members of the next generation, who would be socialized as children within the new institutions, could have a significantly stronger devotion to community welfare. In that case, (f) would be false but (g) would still be true.

Should premise (f)—the claim that reciprocity motivations and limited benevolence have a genetic basis—be granted? The theory of reciprocal altruism only shows how it is *possible* for a gene for reciprocity to be selected; it does not show that, if reciprocity is widespread, this is because it *actually* has a genetic basis. What Philip Kitcher has called "pop sociobiology" makes the jump from possibility to actuality by assuming the truth of "adaptationism" (*op. cit.*, pp. 213–240). Adaptationism is the view that natural selection is the only significant source of evolutionary change, and that every heritable trait of a creature has been fashioned by natural selection so that the trait maximizes the creature's (or its ancestors') inclusive fitness. (Consequently, one job of the biologist is to figure out what function the trait does or did have.) The first claim ignores (or underestimates) the effects of chance and genetic drift, and the second ignores pleiotropy, linkage, allometry, and so on. But, though arguments for (f) that appeal to adaptationism are too a priori, better ones may be available or forthcoming. The existence of reciprocity motivations in birds and baboons—where a biological explanation of their origins is the only plausible one—surely justifies a suspicion that the motivations have a strong genetic basis in us, too.

PSYCHOLOGICAL DETERMINISM

Defenders of the Golden Rule, absolute pacifism, and Christian or utilitarian sainthood should not reject (f) or (g). That means that they must reject (b), (d), or (e). (d) says that anyone who accepts their moral principles will avoid frequent backsliding only if she has the right behavioral dispositions and felt inclinations, and (e) says that she will have these only if her benevolence motivations are very strong. Only then will she have the "energy" necessary to hit her "target."

The target/value judgment and energy/motivation analogy that I have used is certainly misleading insofar as it suggests that urges, felt inclinations, cravings, and behavioral dispositions cannot by themselves have determinate objects. Someone could have a desire to fondle a particular body or a craving for a particular brand of chocolate; it is obviously not true that any sweets will satisfy a sweet tooth or any body a sexual urge. But the analogy might also be challenged, because it suggests that motivation is distinct from valuation and that one can hit one's target/value judgment only if one has energy/mo-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tivation of the requisite kind and strength. In fact, most defenders of the moral principles we are considering will reject the analogy as flawed for just this reason. In doing this, they reject (d), what I have called "psychological determinism."

Psychological determinism, as I understand the view, says, first, that one's sincere acceptance of a value judgment has causes and is not under the direct or immediate control of one's will; second, that one's felt inclinations and behavioral dispositions also have causes; their existence and strength are not under one's immediate control; third, that highly valuing a state of affairs is not the same thing as being strongly motivated to bring it about; one can have the former without the latter, in which case one will suffer "weakness of will"; and, fourth, weakness of will in such a case is unavoidable, because motivation of the requisite kind and strength is causally necessary for any action required by one's value judgments. Notice that the third claim is that high valuation is consistent with lack of strong motivation, not that it is consistent with no motivation at all. If someone claims to place great value on something but makes no attempt to secure it when she has the opportunity to and she feels no regret or guilt subsequently, we would not say that she fell prey to weakness of will; we would say that she does not really value what she claims to. Unless one has *some* disposition to seek x and *some* felt inclination toward it, it would not be the case that one really values or approves of x. But the possibility of weakness of will exists only because one *can* value something quite highly even though one's disposition to seek it and felt inclination for it are weak (though still strong enough to justify the claim that it is a case of weakness of will rather than a case of not really valuing what one claims to).[13]

As regards the first claim, the psychological determinist holds that one's value judgments and method of reasoning about evaluative questions are often affected by nonrational factors—in particular, by the behavioral dispositions and felt inclinations that evolution, conditioning, and socialization have put into one's character. Sociobiology can contribute something to an account of the *origins* of moral beliefs only if this first claim is correct. If a person's genes causally influence which moral principles she will come to accept, they do so indirectly, by causally influencing the nature and strength

[13] Here I follow Gary Watson ["Skepticism about Weakness of Will," *Philosophical Review*, LXXXVI (1977): 316–339] in thinking that "weakness of will" consists in highly valuing a state of affairs that one is only weakly motivated to bring about. For a fuller discussion of the "evaluational" and "motivational" components of desiring, see Watson's "Free Agency," this JOURNAL, LXXII, 8 (April 24, 1975): 205–220. See also Michael Stocker, "Desiring the Bad: An Essay in Moral Psychology," this JOURNAL, LXXVI, 12 (December 1979): 738–753, on the same theme.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of her motivational energies. This point is uncontroversial, I think, and would be accepted by sociobiologists and their critics. If the incest taboo has a genetic basis, it is not because there is a gene for the belief that sex with close relatives is bad, but rather because there is a gene for aversive feelings to sex with close relatives, and people with such feelings tend to form the value judgment that incest is bad. If it were not for the psychological mechanisms (whatever they are) in virtue of which the feelings tend to produce the value judgment, a sociobiological explanation of the incest taboo would be impossible. What are the mechanisms and how do they operate? These are questions for the psychologist rather than the sociobiologist. If the psychological determinist's fourth claim is correct and one can hit the target posed by one's value judgments only if one has the requisite motivational energies, then perhaps the mechanism in question is essentially a dissonance reducer. People do not like to miss whatever target at which they aim, so perhaps they unconsciously make sure that they aim only at what they can hit; perhaps they tend to adjust their values to their motivations. Sociobiologists tell us that there is a genetic basis for reciprocity and the retributive emotions (to feel gratitude toward those who help us, resentment and vindictiveness toward those who harm us). If this claim is correct, *and* if something like this mechanism exists, *then* we can have a sociobiological explanation for why so many people in diverse cultures and historical epochs have believed that free riding is wrong, wrongdoers deserve to be punished, and everyone has a right to self-defense.

The fourth claim made by psychological deterministists is that one can act in accordance with a certain principle or judgment, do the thing of which one on reflection approves, only if one has dispositions and inclinations of the appropriate kind and strength. Value judgments by themselves are not sufficient to cause the action prescribed by the judgments; the conviction that turning the other cheek is a moral duty will not by itself cause or enable one to turn the other cheek. This claim does not deny value judgments *any* causal role in action. In particular, it does not turn value judgments into mere epiphenomena of one's behavioral dispositions and felt inclinations. One reason why that would be an obvious mistake is that most moral judgments are "underdetermined" by the dispositions and felt inclinations that the psychological determinist claims are causally necessary for the behavior that the judgments enjoin. Even if there are tight limits (imposed by our biology or upbringing or whatever) on the nature and strength of moral motivation, there are several incompatible moral principles which a person with any one kind of motivation could accept and to which she could conform. This point can be made by means of the value judgment/target and motiva-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tion/energy analogy. The strength of the archer, how far she can pull back the bow string, sets a limit on which targets she can hit; but it does not explain why she aims at one target rather than others that lie within range. Consider again reciprocity, the retributive emotions, and the moral principles that people so motivated tend to accept (that free riding is wrong, wrongdoers deserve to be punished, and everyone has a right to self-defense). My present point is that these principles are vague; many specific moral principles that are incompatible with each other are compatible with them. There are several competing views about what doing one's "fair share" means. The "eye for an eye," strict liability view of punishment in the Old Testament and the *mens rea* view of guilt in Christianity and Kantian moral theory are competing forms of retributivism. And there are disagreements about the scope of a right to self-defense (e.g., whether it entitles the mother to kill the fetus, an "innocent threat" to her life, in the "craniotomy" case). This means that the retributive emotions are not the sole cause of the different versions of these principles that different people accept. A full explanation why individuals hold the particular moral beliefs that they do will cite the role that their upbringing, culture, and class interests also have in shaping their considered moral judgments, and it will cite the role that rational thought, the desire for logical consistency, and the desire for moral principles that sit well with one's empirical and metaphysical beliefs all play in transforming a motley collection of considered moral judgments into something approximating a moral theory. The belief that wrongdoers deserve to suffer has causal antecedents, just as the tendency to feel anger toward and to strike out against those who have harmed one has causal antecedents. But there can be no doubt that the judgment is not the same thing as the felt inclinations and behavioral dispositions; and granting that someone's having the motivations is one cause of her acceptance of the "retributivist" moral judgments, there can also be no doubt that her acceptance of the judgments has many additional causes. One can agree with the fourth claim made by the psychological determinist and still reject Wilson's crude idea that a person's considered moral judgments are just the epiphenomena of emotional responses emanating from the hypothalamic-limbic system (*op. cit.*, p. 563).

To return now to the argument, I think that psychological determinism is the premise of the argument which most defenders of principles enjoining sainthood or pacifism will wish to attack. In place of psychological determinism's fourth claim, they will embrace a libertarianism that insists that each of us has a causally unencumbered will. It denies that one's motivations (or any other features of one's character) constrain which choices one can make or which actions one can successfully perform. It holds that, whether or not

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

one succeeds in living up to one's principles depends on how strong one's will is, not how strong one's inclinations and dispositions are. What is more, strength of will has no causal determinants, genetic or environmental.

A defense of psychological determinism falls outside the scope of this paper. I shall only state that I find the view compelling and that, in my view, if a defense of pacifism or sainthood depends on a libertarianism that is inconsistent with the view, then the defense is in deep trouble. As I see it, defenders of such principles are better advised to grant (d) and argue against (b), the "ought implies can" principle, or (e), the claim that only a strongly benevolent person can be a saint or pacifist.

One can accept psychological determinism but reject (e) if one can identify another source of the motivational energy necessary for sainthood (besides strong benevolence). It might be suggested that a "desire to be moral," in someone who believed that she had a duty to be a saint, could motivate saintly behavior. But if the desire is conceived as an all-purpose motivational energy that will allow one to live up to one's moral principles *whatever* they happen to be, it is doubtful that such a desire exists.[14] The suggestion that action could be motivated by this desire is more plausibly interpreted as the claim that a person's value judgments, whatever they happen to be, can by themselves motivate the action they enjoin. This claim is of course inconsistent with psychological determinism. A better objection to (e) is that *pure egoism* can motivate saintly behavior if it is conjoined with the belief that there is a God, the belief that God commands saintly behavior, and the belief that people who disobey God's commands are punished in an afterlife with eternal damnation. Some Christians who accept the "turn the other cheek" injunction will have no qualms about motivating adherence to it in this sort of way. Even Singer can save his principle enjoining utilitarian sainthood from refutation by our argument by exploiting this fact about egoism. If human beings are strongly egoistic, and if egoists will behave like saints only if they hold these religious beliefs, then, he can say, people should convince themselves of the truth of the religious beliefs, through nonrational means like brainwashing, if necessary.[15]

[14] W. D. Ross and H. Pritchard assume the existence of such a motivation. Cf. Christine Korsgaard, "Skepticism about Practical Reason," this JOURNAL, LXXXIII, 1 (January 1986), pp. 9–10.

[15] In *The Expanding Circle*, Singer accepts psychological determinism and agrees that strong benevolence is rare for the reasons put forth by sociobiologists. But he does not think that egoism plus religious belief is the only other way of motivating saintly behavior. He argues that [illegible]

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Although this shows that (e) is false, that the argument in which it is a premise is unsound, and hence that Singer's principle is not refuted by the argument, the victory thereby secured for the principle is surely a hollow one. If the price that has to be paid to ensure full compliance with the principle is the sacrifice of our rationality and autonomy, we should seriously doubt whether it is worth it. The principle assumes that the promotion of happiness and prevention of misery are the only things of value, and if autonomy has independent value (as it surely does) then the principle is simply false.

Perhaps there are yet other ways in which saintly behavior can be motivated. Consider people who make large sacrifices of time and money to protect the wilderness from despoliation by real estate developers and strip miners. Surely some of these people value natural beauty for its own sake and want the Grand Canyon to be protected even if no one is around to notice its existence or enjoy its beauty. These people do not believe that the Grand Canyon is itself a sentient being. Their sacrifices for it cannot be motivated by egoism, benevolence, or reciprocity. The existence of such people surely proves that upbringing, socialization, or conditioning somehow can produce the motivational energies necessary for the pursuit of an ideal that has nothing to do with one's own or others' welfare. If this point is admitted, then how can we rule out the possibility that a certain type of upbringing could produce a motivational energy which causes no fellow feeling or affection for others, but which enables one to make the large sacrifices of the utilitarian saint? Perhaps we cannot. But we should bear in mind that the number of people who make large sacrifices for ideals that have no connection to anyone's welfare is small. If the motivation that enables them to make their sacrifices is just as rare as strong benevolence (and just as difficult to acquire if one lacks it), moral principles enjoining saintliness or pacifism are still seriously challenged by the argument we have been considering.

CONCLUSION

Since the only reason why (e) is false is one that does not rescue these principles from refutation by the argument, I conclude that the defenders of these principles must reject (b), the "ought implies can"

---

moral point of view and have a sufficiently strong *desire for consistency* can be saints. I think he is clearly wrong about this. Even if he were right in supposing that people who accept the moral point of view but deny that they have a duty to be saints are somehow inconsistent, it would not follow that if they were consistent they would be able to fulfill the duty. Inconsistency can only occur between judgments or principles, and at most a desire for consistency can motivate one to modify one's judgments or principles. Someone who assents to the judgment that he ought to make the saint's sacrifices but then fails to make them will suffer from weakness of will, but there is no inconsistency in weakness of will.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

principle. They could follow William Sinnott-Armstrong's[16] lead and reject the principle altogether, or they could accept the principle but claim that it is misused in the argument. The second strategy would require arguing that the "ought implies can" principle does not count a lack of motivation causally necessary for the performance of an act as an obstacle or incapacity. If being in two places at once or possessing the strength of a Hercules were necessary to perform an act, then perhaps (*contra* Sinnott-Armstrong) it follows that one has no duty to perform it. But defenders of Christian and utilitarian sainthood might argue that "ought implies can" is true only if the sense of "can" is "nothing prevents her from doing her duty, *if* she wants to—has the motivation causally necessary to—do it." This is the hypothetical "can" of soft determinism. The motivation behind soft determinism is to leave room for judgments of responsibility and positive or negative agent evaluation (e.g., "Hitler was an evil man, deserving of punishment") in a deterministic world. Similarly, the motivation for a hypothetical reading of the "can" in "ought implies can" would be to create room for the truth of a judgment of act evaluation like "everyone has a duty to do x" in a world in which psychological determinism is true and the vast majority lack the motivational energy causally necessary to do x.[17]

I suspect that whether or not "ought implies can" is true when the can is a categorical one all depends on whether a "realist" meta-ethics that takes the truth conditions of moral judgments to be quite independent of our cognitive and motivational capabilities is superior to a "constructivist" meta-ethics that insists that we only have reason to accept and act on a moral code that is adapted to us—our needs, interests, abilities, and so forth. This is a large question. It is at least as large and no less philosophical than the question whether psychological determinism is true. It is noteworthy that the soundness of an argument against certain moral principles which was suggested by sociobiological speculations about the genetic basis of egoism and altruism ends up depending not on whether there is sufficient empirical evidence to buttress the speculations, but rather on two premises, psychological determinism and "ought implies can," that seem more "philosophical" than "empirical." This can only be a disappointment to those like Wilson who had hoped that sociobiology would put ethics on a more scientific footing.

DANNY SCOCCIA
New Mexico State University

[16] "'Ought' Conversationally Implies 'Can'," *Philosophical Review*, XLIII (April 1984): 249–261.

[17] Derek Parfit [*Reasons and Persons* (New York: Oxford, 1984), p. 15] and Robert Merrihew Adams ["Motive Utilitarianism," this JOURNAL, LXXIII, 14 (August 12, 1976): pp. 467–481] seem to accept an "ought implies can" principle with a hypothetical "can."

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

COMMODITIES, LANGUAGE, AND DESIRE*

A central topic for economic science is people's desires for goods of different sorts and how these desires lead to exchange. These matters are modeled in the standard theory in terms of a finite set of "commodities." But as to the nature of commodities and how they are individuated, the standard theory maintains an aloof silence. In this paper, I offer an account of commodities that breaks this silence. I take commodities to be goods in the trading of which language enters in a certain way. The critical condition for there to be such trading is, I shall say, the semantic power of natural language. The gains from trade in commodities are therefore not to be attributed to the actions of autarchic individuals, as liberal individualism maintains. They rest upon traders' use of resources that cannot but be held in common.

The traditional account of commodities, that of general-equilibrium theory, holds that commodities are n kinds into which fall the things that consumers desire and acquire through exchange. It is usually assumed that consumers know that all tokens of the ith commodity ($i = 1, \ldots, n$) are alike in terms of their desirable, and their detestable, attributes. This leads to the "law of one price," and thence to the great theorems on the optimality of competitive economies. Large questions are here passed over: Why should desire be in kinds? Why should trade? What determines the typology; why not coarser, or finer, or incongruous? Why kinds homogeneous to desire? Whence the consumer's knowledge of the properties of the token of the commodity i she would acquire if she bought? Why, for that matter, does she think that what she would get would *be* an i? The notion of commodity that I shall present throws light on these questions.

The causal explanation of trade in types begins with the psychological fact of the notionality of our desires. This opens the way for a massive economy of resources in how we conduct exchange; all that a prospective buyer needs to know, all that a would-be seller needs to convey, is the type or types to which a prospective object of trade belongs. There emerges trade by verbal classification of what is offered and sought. It is classes thus verbally picked out which I shall call *commodities*. Trade of this sort may span great distances at low

* I am grateful for helpful discussions and criticisms to many, especially to John Broome, Robin Cubitt, Susan Hurley, Isaac Levi, David Owens, David Papineau, Philip Pettit, Hyun Song Shin, and Holly Smith, and most of all to Elizabeth Fricker.

0022-362X/90/8707/346–368 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

cost. It has been facilitated by many developments, some concerning things, some signs: by technical progress both in the shipment of goods and in the transmission of symbols, by the growth of methods of mass production, and by the development of a new law of trade, which protects the buyer by obliging him who offers to supply 'a ϕ' to supply something properly so described. But the fundamental conditions of its existence are the logical form of our ordinary wants and the information content of our ordinary predicates.

In section I, I suggest that a theory of consumers' behavior should take consumers to be moved by notional desires. In section II, I define commodities and contrast trade in commodities with its opposite, "medieval" trade. In section III, I note—*pace* general-equilibrium theory—the heterogeneity of traded types with respect to their "graders"—the attributes that consumers value or abhor in things of a given type. In section IV, I connect two sets of beliefs: those of consumers about the graders of that which they would acquire if they engaged in commodity trade in a type F; and those which make up the "everyday theory" of Fs and so fix the meaning of the predicate ϕ that denotes F. Section V establishes, in a simple model in which Fs can be traded in an economy E either as a commodity or medievally, the following claim: the type F will be a commodity in E just if the everyday predicate ϕ bears enough information, in a certain exact sense, relatively to its syntactic complexity.

I. WHAT CONSUMERS WANT

It is beyond doubt that wants to acquire things are predominantly wants to acquire things as instantiators of general concepts. To be more precise, the wants of a typical consumer C are on the whole wants that we may describe by Quinean notional-desire sentences of the form

$$D(\phi)$$

$$\mathcal{D}(\exists x)(\phi x \wedge \eta x)$$

where ϕ represents a predicate that denotes a type F of thing, $\mathcal{D}$ the operator *C desires that*, and η the predicate *will be had by C*. I shall call such sentences as $D(\phi)$ *D-sentences*. According to $D(\phi)$, C wants whatever satisfies a certain description (ϕ). She wants an F qua F; she wants "relief from F-lessness." Such desire ascriptions have three familiar features: they are nonextensional; they do not imply the existence of a token of the type; and they are of an "unfussy" desire —a desire for *any* token of the type.[1]

The postulate that consumers' desires are notional desires has

[1] The form $D(\phi)$ might be used to ascribe certain singular desires if we admitted as substitution instances of 'ϕ' indexical phrases, such as 'this jersey, returned dry cleaned'. I exclude them.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

theoretical advantages. It helps explain the existence of "commodities." And, into the bargain, it provides a convincing explication of the consumer's desires of received theory, which there appear as unanalyzed primitives, and so strengthens the foundations of that theory. The basic sentences of that theory, in its Hicksian version,[2] may be paraphrased thus: "The desirability for C of getting x units of commodity i and y units of commodity j is no less than that of getting x' units of i and y' units of j." The theory contains no sentence that asserts or implies the existence of these units; on the contrary, preferences are defined on the whole of a certain Euclidean space, while only part of it represents producible "baskets" of commodities.[3]

The sentence $D(\phi)$ makes explicit the logical form of consumers' desires, and that is also to the good,[4] but in other respects D-sentences give a rudimentary model of them. While $D(\phi)$ describes a desire of unspecified strength for an unspecified number of a single type of thing, a fully fledged theory deals with desires for specified quantities of several types of thing—desires for "baskets" of goods —and orders these desires by strength. Not all these underspecifications matter for my project; those which do I shall make good in due course. Since $D(\phi)$ lends itself to the interpretation that C desires true a certain proposition, a decision theory in the style of Richard Jeffrey[5] provides a natural home for an enriched version of $D(\phi)$ which allows for degrees of desire; and that is where I shall in due course house mine.

Not all wants are "notional" wants, describable by D-sentences. A person may want a particular object K in a sense that both entails that K exists,[6] and means that there is no presumption that the want can be satisfied by any other object, however qualitatively similar to K. One may have a want for *that* doggy in the window or for a *certain* beloved jersey back from the cleaners. Does a theory of exchange that has only D-sentences or their like to describe consumers' wants therefore lack essential resources? I think not, for two reasons.

First, casual observations of consumers' behavior, and of its every-

[2] J. Hicks, *A Revision of Demand Theory* (New York: Oxford, 1956).

[3] Further evidence for the notional reading of standard theory is its transitivity postulate. 'If C prefers an F to a G and a G to an H then she prefers an F to an H' is a silly postulate if 'prefers a' is read relationally.

[4] Of economic theory. The form $D(\phi)$ highlights their nonextensionality and so alerts against fallacious reasoning. The numerical typology of commodities leaves it unsettled whether 'i' denotes a concept of the consumer, or of the supplier, or a set of actual objects. This ill-equips the theory to deal with such possible facts as that all and only white bread is carcinogenic bread.

[5] *The Logic of Decision* (Chicago: University Press, 1965).

[6] See G. Evans, *Varieties of Reference* (New York: Oxford, 1982) on such "genuinely singular thoughts."

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

day explanation, suggests that exchange-motivating consumers' wants are typically notional. To 'What's he gone to the shops for?' we reply 'He wanted to get a 13 amp plug'. Surely, not a particular 13 amp plug? Second, it is arguable that the desires of the fully rational consumer must always be notional—even when it takes perception of an actual object to cause them; that a consumer fails of rationality if there is no set of attributes such that were the trusted shopkeeper to assure her that there is another thing in the back of the shop which has them all, she would still insist on the one in the window. This view about rationality may be latent in general-equilibrium theory and may have encouraged that theory to treat, as it does, only of notional desires.

II. COMMODITIES

1. The Definition of a Commodity. The general notion of a "commodity" I propose is of *a type of thing traded by means of a description.* More fully, I shall call a type F a "commodity" in an economy, E, if it sometimes happens that one member of E makes an offer to supply, subject to some condition, what he describes as being a unit of F; another member accepts the offer; and the agreement is carried through. An "economy" will be defined, implicitly, as a speech community whose members trade with each other.

By an $O(\phi)$-sentence I shall mean a completed sentence of the form

'I, the undernamed, will supply with one ϕ any person who agrees to meet the following condition: . . .'

By a *ϕ-offer in the economy* E, I mean an offer made by one member of a group E to another by uttering an $O(\phi)$-sentence, or an equivalent form of words,[7] in the language of E. A *ϕ-exchange in* E is an episode that consists of a ϕ-offer in E followed by its acceptance and consummation. I shall call the offerer the *supplier,* and the recipient of the offer the *consumer.* I define F to be a *commodity in* E if and only if

ϕ denotes the type F in the language of E, and there are ϕ-exchanges in E

(and then, there is *commodity trade in Fs in E*). In brief, F is a commodity somewhere if people there offer tokens of F in so many words and others take up these offers without more ado.

My definition of commodity needs some glosses. It says nothing of the content of the condition in an $O(\phi)$-sentence. Typically the con-

[7] It would be a pity to make the notion of commodity hinge on inessential variations in the wording of offers. It is not a trivial task to characterize the appropriate notion of equivalence. The two essential things are sameness of information content, and nonredundant use of the predicate ϕ.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

dition will stipulate two things: what if anything the consumer must do to take possession of the (token of) F, and that the consumer supply something else, some *quid pro quo,* in exchange. I shall henceforth suppose, for harmless simplicity, that she need take no active steps to collect: the F is brought to the doorstep. And that there is a *quid pro quo:* this may well be, but need not be, a monetary payment. Nor is my definition specific as to the volume in which Fs must be traded for F to be a commodity: one F will do.[8]

I take the language of E to be one in which all members of E are competent and know each other to be; more specifically, I take it to be a nonspecialized portion of some natural language (here, for harmless convenience, English), and ϕ to be therefore an everyday predicate of the latter.[9] But I have put no restriction on how elaborate the description contained in ϕ may be: ϕ may be anything from a one-word brand name to an arbitrarily complex predicate pages in length.

Makings of ϕ-offers are acts of "notional supply." There is no particular F that the offerer commits himself to supplying: ϕ-offers are noncommital just as $D(\phi)$-ascribable desires are unfussy. They "fit" such desires without wasting words: nothing is promised which is not present in the desire. Herein lies part of the efficiency of the social contrivance of commodity trade, and part of the secret of its success.

2. *Medieval Trade.* Commodity trade as I have defined it is only one way in which people might, and do, trade in things. It is found, in approximated form, in trade through mail order and the small ad, and in those temples of trade in which the buyer may study detailed descriptions of lawnmowers and fitted sheets in a shiny volume, then to see, no sooner is the decision for an F made, emerging from the clockwork, the one that all unknowingly she has made *her F.*

There is another way of trading, in sharp contrast, in which the consumer agrees to acquire (on some condition) a particular thing with which she is acquainted. The acquaintance is more than a nodding one. The consumer, before agreeing, has been able to inspect the object in various ways. Such inspection characterized the bulk of medieval trade in consumer goods. One had a good look round a

[8] This openness allows definite and other uniquely instantiated descriptions to define commodities; if it is a little at odds with the presystematic notion, it is harmlessly so.

[9] My account of commodities could be generalized to allow arcane technical predicates such as '22 mm. obtuse elbow'; but this would complicate the analysis of section IV, with little gained. Although such predicates have a role in trade in consumer goods, it is a walk-on one.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

horse, just as one often can, today, have one round what today does the work of the horse. I shall call *medieval trade* trade with such prior inspection of a particular object offered in exchange.

Commodity trade is trade *with description and no inspection.* Medieval trade is trade *with inspection.* If in addition there is *no* description, we may speak of "strictly medieval" trade. But this is a heroically ideal case, for some element of description is almost inevitable in any exchange. The seller must normally commit himself to the claim that the offering is his to sell. Even if the goods are to be inspected, he frequently initiates his interaction with customers by broadcasting that he has wares of a certain kind for sale. Where he does not say so, his actions may nevertheless bespeak so. Imagine street-market trade in unlabeled apples—here apparently we have strictly medieval trade. But a court might recognize that the customer has a right of redress when they turn out rotten; and she can hardly have one unless they have been implicitly offered as apples. It may well be that they were so offered and so accepted. Indeed, we may guess that the displaying of the apples together in their rough basket on a dusty patch is motivated by the seller's expectation that the passing consumer will see them as apples for sale, not as an assemblage of objects otherwise conceived by him, and that she will do so by recognizing that in displaying them in the basket the seller intended to produce this effect on her.[10]

In the remainder of this paper, I shall be concerned with the contrast between commodity trade and medieval trade *tout court;* between a sort of trade with description and no inspection, and a sort of trade with inspection[11] not necessarily to the exclusion of description. But these do not exhaust the modes of trade.[12]

III. HETEROGENEITY

One purpose of this paper is to explain the existence of commodity trade. Commodity trade is inspectionless trade in a type F, barely so described. An explanation must therefore come to terms with a fundamental feature of, and potential obstacle to, such trade: there are Fs and Fs. To this heterogeneity, and the way in which it affects the attractiveness of ϕ-offers, I now turn.

[10] If so, the putting out of the apples was an utterance with the S-meaning that he has apples for sale: see H. P. Grice, "Meaning," *Philosophical Review,* LXVI, 3 (July 1957): 377–388.

[11] Inspection opportunities are more or less constricted. They range from peering at one side of a pork chop through a cellophane wrapper to taking a consumer durable home for a trial period.

[12] There is trade with inspection that is not medieval, as when, for example, one is made a notional offer and may inspect a sample from the type. There is trade with description and no inspection that is not commodity trade, as when the parties bargain, exchanging offer and counter offer.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

1. The Law of Trade. The history of the law governing the exchange of goods attests to the rise of commodity trade. In a system in which the buyer saw the goods and could examine them—and in which the proximity of trade meant, furthermore, that she often knew the supplier himself—it was easy to maintain that the buyer had only herself to blame if what she bought was defective. Moreover, putting the onus on the buyer may have favored trade, for, since a buyer knows just what she is looking for, it made for a more efficient allocation of epistemic labor. By contrast, when trade is by description and without inspection, *caveat emptor* encourages negligence and dishonesty in the description, which give trade a bad name. If law is the hand-maiden of trade, we should therefore expect the rise of commodity trade to produce a doctrinal shift, away from *caveat emptor* and toward a doctrine that upholds the duty of the supplier to supply what he says he will supply. This is just what took place in the English law, both in common-law judgments and in a sequence of statutes.[13]

Common-law judgments about the proper description of wares are of interest to us for another reason, too. They embody an account of the information that is conveyed by ascribing an offered good to a heterogeneous type. It will be the basis for mine.

In 1815, Ellenborough (Gardiner versus Gray) finds against a supplier of a batch of moth-eaten cloth on the grounds that "a man does not purchase worsted to lay it on a dunghill." In other judgments it is held that a sufficient accumulation of defects in an article constitutes a breach of contract. In 1956, Birkett [Karsales (Harrow) Ltd. versus Wallis] finds for the plaintiff against the supplier of a second-hand "car" on the ground *inter alia* that "a car without wheels is no car at all."

Collectively, these judgments, and corresponding statute, appeal to a number of principles: first, that he who contracts to supply an F, so described in a verbal formula such as 'motor car', has a duty to supply something with a sufficiency of certain ancillary properties $G_1, G_2, \ldots$ (for cars, the wheeled property, the starts-up property, . . . ; for cloth, the holeless property, the keeps-warm property, . . .). Second, $G_1, G_2, \ldots$ are the properties that make the thing fit for the purposes for which people normally want Fs ("merchantable" as an F). Third, they are properties that it is "reasonable to expect" in Fs. Fourth, correct description as an F requires what duty requires—that the object possess a sufficiency of the Gs.

[13] See G. Borrie and A. L. Diamond, *The Consumer, Society and the Law* (Harmondsworth: Penguin, 1980, 4th ed.).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

2. Grading Properties. Unlike the "commodities" of traditional theory, commodities in my sense are in general *not* homogeneous. One token of the commodity F may differ markedly from another, equally an F, and the differences may be ones that matter to consumers.

It is convenient to introduce the notion of a "grading property." Henceforth I shall denote by FG the conjunctive type of things that are both (of type) F and (of type) G, and by $F\bar{G}$ the type which are F and not G. I shall say that G is a *grading property* or *grader* of F *for* C if C discriminates in her desires between having an FG and having an $F\bar{G}$; that is, if she has a preference one way or the other. If the preference is in favor of the former, I call G a *positive grader* of F *for* C. If G is a grader of F, I shall also say that the type FG is a *grade* of F.[14]

Graders abound. Being brown is a grader of eggs (for me), being salmonella-free a positive grader of them (no doubt for all); being moth-eaten is a (negative) grader of cloth; having wheels is a positive grader of motor cars, if we may allow, *pace* Birkett, the notion of one without them.[15]

Intuitively it is clear that the desirability of accepting a ϕ-offer depends on your expectations about how good an F you would get (if any) if you did. That is, on your assessment of your chances of getting an F; and of the chances, if you do, of its being of various grades. The desirability of accepting a ϕ-offer is a key magnitude in the theory of commodities. The factors determining it are complex. It is therefore expedient to focus on a special case. I shall call this case *pure* commodity trade.

3. Pure Commodities. Your confidence that accepting the ϕ-offer would bring you an F depends in the first place upon your *understanding* of the offer and your belief in the supplier's understanding of it. My definition of commodity makes sure that this condition is met. It depends in the second place on your *trust in the supplier*. By this I mean your belief that he is both sincere and competent with regard to carrying out his undertaking. Doubts on these fronts are possible enough, but I exclude them from the case of pure com-

[14] Nondichotomous attributes that matter to people, such as numerical ones, can be represented arbitrarily well by a set of graders.

[15] What it is that makes property G grade F positively is a question that lies outside my theory. This is where K. J. Lancaster's "characteristics" come in ["Change and Innovation in the Technology of Consumption," *American Economic Review*, LVI, 2 (May 1966): 14–23]. Among Fs, those which are Gs are the more desirable the more that G-ness (given F-ness) contributes to C_1-ness, C_2-ness, etc., where C_1, C_2, . . . are desirable characteristics.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

modity trade. In pure commodity trade, then, the consumer is certain that, were she to accept the offer, she would get a thing properly described as an F. I assume further, in the pure case, that the consumer has no self-doubts: she is clear that, were she to accept, she too would do her part and deliver the *quid pro quo*.

C's view of how likely she would be to get a G, given that she would get an F, is a matter of her background beliefs together with whatever clues her particular ϕ-offer provides as to the grade of the F she would get. These indications are of two kinds. First, there is the information that what is offered is an F, and whatever may be inferred therefrom. Second, there may be in addition "incidental" information about the offering generated by other constituents of the content of the ϕ-offer, or by its context—the name of the supplier, the price, the glossiness of the sales blurb, the ideological position of the newspaper in which the offer appears. In the case of pure commodity trade, I exclude such incidental information.

All this may be made precise within a simple causal version of decision theory[16] which will also serve us later. For any proposition a, write $Pr(a)$, $Ut(a)$ for C's degree of belief that a and degree of desire that a. The fundamental equation of the decision theory expresses C's degree of desire for a possible action of hers in terms of her desires for it conjoined with alternative consequences of it, consequences that make up some partition of possible worlds. Letting a describe the action, $\{b_1, \ldots, b_n\}$ the consequence partition, and $\Box\!\!\rightarrow$ indicate the subjunctive conditional, the equation is:

$$(1) \qquad Ut(a) = Ut(ab_1)Pr(a \Box\!\!\rightarrow b_1) + \cdots + Ut(ab_n)Pr(a \Box\!\!\rightarrow b_n)$$

As the notation suggests, (1) is a form of the expected-utility principle, and I shall also speak of the values of Ut and Pr as C's utilities and subjective probabilities.

The definition of a pure commodity involves only the belief function Pr. Suppose F is traded as a commodity, and consider an arbitrary consumer C who receives an arbitrary ϕ-offer O. Let 'Pr' denote C's belief function at a moment just before the receipt of O. Let us assume for simplicity that F has for C a single positive grader, G (the generalization is straightforward). Write E for the evidence set with which C is provided by the offer O. E will include the description of the offering as an F, the name of the supplier, the nature of the *quid pro quo*, and features of the occasion. Write a for the proposi-

[16] See, e.g., A. Gibbard and W. Harper, "Two Kinds of Expected Utility," in *Foundations and Applications of Decision Theory, Vol. 1*, C. A. Hooker, J. J. Leach, and E. F. McClennen, eds. (Dordrecht: Reidel, 1978).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tion that *C accepts the offer*, q for *C fulfills the quid pro quo*, f for *C acquires one thing, an F*, and g for *C acquires one thing, a G*. Then O is a *pure ϕ-offer (for C)* if the following two conditions hold:

$$Pr(a \,\Box\!\!\rightarrow f \mid E) = Pr(a \,\Box\!\!\rightarrow q \mid E) = 1 \tag{2}$$

$$Pr(a \,\Box\!\!\rightarrow fg \mid E) = Pr(a \,\Box\!\!\rightarrow fg \mid f) \tag{3}$$

I shall say that F is a *pure commodity in E* if it is a commodity in E and all ϕ-offers in E are pure. The *condition of trust*, (2), says that C is sure, when her pre-offer beliefs have been updated for the evidence of the offer, that if she accepted the agreed exchange would go through; (3), the *condition of sufficiency*, that, given her background beliefs, all the information the offer provides about the grade of F of the thing she would get if she accepted[17] is contained in the information that she would get an F.

If F is a pure commodity then, typically, the desirability of having an F cannot depend on who supplies it, on the occasion of the ϕ-offer, or on the exact price you pay.[18] None of these features, nor any that may be inferred from them, grades Fs. All this accords with the assumptions of received consumer theory.

4. Commodity Trade and D-Sentences. The occurrence of commodity trade in Fs in an economy, and the prevalence there of notional desires for Fs, go together; ϕ-offers and notional F-desires are made for each other. For what explains a consumer's acceptance of a pure ϕ-offer? Since she believes that $a \,\Box\!\!\rightarrow f$, she has a motive for acceptance if she has a desire whose satisfaction condition is f. But f is precisely the satisfaction condition of a notional desire for an F; of a desire expressed by the sentence $D(\phi)$. Thus, if she has such a notional desire, then, provided it is sufficiently strong, her action is explained.[19]

[17] Note the narrow scope of the quantifier in the conditionals; e.g., $a \,\Box\!\!\rightarrow fg$ does not say that there is an FG that C would acquire if she accepted.

[18] The desirability of having, not of the decision to buy! If F is pure, then all you learn *ex ante* that you care about in Fs, about the one you would get, you learn just from knowing it will be an F. If you cannot infer its exact price, its provenance, just from this, then if F is pure these properties do not grade it. But an F's desirability *may* depend on these properties, consistently with purity, if the description of it as 'a ϕ' has implications for them, as it may: if someone knows that seats in the stalls usually cost £15–50, or that most shoes come from Varese, the purity of these commodities is unsullied by the infiltration of these features into the texture of her desire for the kind itself.

[19] The supplier does not in general know who the recipients of his message are. Yet he intends, like a newsreader, to address whomever these may be. This multiple speech-act feature of ϕ-offers means that the indexical ϕs mentioned in fn. 1 may also support commodity trade.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Per contra, nothing but a notional F-desire provides her in general with a motive for acceptance. In particular, a singular desire does not; for a ϕ-offer provides her in general with at most the flimsiest grounds for expecting that her acquisition would be the desired object, even where she believes this to be an F. It is medieval trade that singular desires are cut out for explaining.[20]

5. Heterogeneity and Commodity Trade. If the type F is heterogeneous (in terms of its graders), then the desirability of a ϕ-offer depends, evidently, on your conditional probabilities for getting different grades, given acceptance. In the next section, I shall discuss the source of these conditional probabilities for typical consumers. If those for positive grades are low, so, *ceteris paribus,* is the desirability of a ϕ-offer. But it is low only given price. If price is free to move, we cannot conclude from its lowness that the prospects for commodity trade in Fs are poor.

There is, however, another effect of heterogeneity that can only militate against commodity trade. How *sharp* a picture C has of her prospective acquisition, in terms of its graders, depends on *how much information* f contains about them. Little grade information in f, and a vague picture, threaten commodity trade in Fs. For one thing, risk aversion diminishes the absolute desirability of acceptance. But also, as we shall see in section V, a vague picture increases the *relative* attractions of medieval trade.

Heterogeneity of F works against commodity trade in F. But if commodity trade in Fs there is, all these Fs exchange (in equilibrium) at the same price. In the present theory of commodity trade, the law of one price is thus re-established, liberated from the traditional assumption of product homogeneity. In equilibrium, price must be uniform, because all pure ϕ-offers are equally desirable in respect of all features *other than* price. In particular, that which is promised and expected is always the same—'some ϕ'.[21]

[20] Medieval trade may also cause them. And it may be designed to; the seller acquaints us with the doggy, we fall in love.

[21] The law holds too in medieval trade in "opaque" heterogeneous goods, such as Akerlof's second-hand cars which, for all that inspection reveals, may or may not be "lemons" [G. A. Akerlof, "The Market for 'Lemons': Qualitative Uncertainty and the Market Mechanism," *Quarterly Journal of Economics,* LXXXIV, 3 (August 1970): 488–500]. In both cases, all decisions to buy give the same uncertain prospect. But the source of uncertainty differs: indeterminacy of perceptible signs in the one, of verbal description in the other. So do the bearers of uncertain properties: in the one, an actual object; in the other, an intensional object—that which the consumer would acquire.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

IV. LANGUAGE AND BELIEF

1. The Shared-Theory Theory. Let ϕ be a predicate of the language of the economy E which denotes the type F. When a member of E believes of a thing that it is 'a ϕ', she not only believes that it is an F but also, crucially, she believes (perhaps only implicitly) to apply to it numerous generalizations describing tendencies of Fs. Thus, if she comes to believe that it is 'an egg', then—barring counter indications—she acquires such beliefs as that it was laid by a hen, would *almost certainly* break if she dropped it, *might* be brown. If she does not acquire a sufficiency of commonplace, tendential beliefs of this sort, she does not understand the sentence 'it's an egg' and she cannot be correctly ascribed the belief that the thing is an egg.[22] I shall call the network of propositions that embed the concept F and inform in this way competent speakers' beliefs about what they believe to be Fs, the *everyday theory* of Fs.

Among the members of a single-language community, meanings, and so the network of beliefs which fixes them, cannot vary much. There is a core of tendential beliefs such that anyone who believes 'it's a ϕ' of an object acquires most or all of them of it. Furthermore, the communicative functions of language are served only because everyone *believes* (at least implicitly) that other members believing the sentence possess them. (If there are higher-order beliefs, too, all the better.) In sum, in a community with a successful linguistic practice in which ϕ denotes the type F, there is iterated belief that, when someone believes 'it's a ϕ', she has beliefs about the object conforming roughly to the everyday theory of Fs.

To have purchase in my model of trade, these considerations about language and belief need some idealizing. I shall take the generalizations of everyday theory to the effect that things of one kind tend, to greater or lesser degrees, to be of certain others to be judgments of numerical *probability*. I suppose the everyday theory of Fs to contain tenets of the form: *the probability that a thing is G given that it is F is k*.[23] I shall write $Pr^0(G/F) = k$ for this, and call

[22] Cf. Donald Davidson, *Inquiries into Truth and Interpretation* (New York: Oxford, 1984). A refined account allows for a division of linguistic labor under which concepts are individuated by *inter alia* a corpus of *none*veryday theory, and understanding requires sensitivity to what such theory says, but not knowledge of it. Such sensitivity is recognized in legal opinion regarding what should be understood by 'a ϕ'.

[23] Everyday theory is definite about some tendencies and wishy-washy about others. The generalizations of everyday theory have, accordingly, two dimensions of strength: strength of the tendency for Fs to be Gs, and the firmness with which they are believed. My idealization suppresses the second of these.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

such tenets *everyday generalizations*. What is the best regimentation of everyday generalizations is a complex question which need not detain us here.[24] I shall call the *everyday profile* of F the set of the probabilities $Pr^0(G/F)$ got by letting G range over all the properties of Fs dealt with in everyday generalizations. Everyday F-theory is partial; it does not treat of all properties of Fs. I shall assume that we do find in it, not only a range of consumers' grading properties of Fs, but also certain others. These are the "trading properties" of Fs that concern the modes and terms of trade in Fs that prevail in the community, e.g., that one gets them in F-eries, or that they cost thousands.

The theory of a shared probabilistic theory of the world which I have been sketching has implications for social science which go well beyond the use to which I shall put it here. It provides a basis, it seems unnoticed, for a version of the so-called Harsanyi doctrine—the claim that *rational* agents possess a common, iteratively known, "prior" subjective probability distribution over all states of the world.[25] The prior provided by everyday theory may be less than comprehensive—gappier, indeed, than Ellenborough's cloth; but the rationality required for its possession is modest and universal: no more than is required to know one's mother tongue.

2. *Consumers' Beliefs About Fs.* I abbreviate $a \,\Box\!\!\rightarrow f$ and $a \,\Box\!\!\rightarrow g$ by f_α and g_α, respectively. We have seen that the consumer's evaluation of a pure ϕ-offer involves the conditional probability she attaches to what she would get being G, given that it would be F, that is, on the magnitude $Pr(g_\alpha|f_\alpha)$. In the model of trade of the next section, I shall take these magnitudes to be just the corresponding probabilities of the everyday profile of F, that is, I shall set, for all F, G, C

$$Pr(g_\alpha|f_\alpha) = Pr^0(G/F) \tag{4}$$

(4) follows from two assumptions. First, the consumer is a layperson in F matters. This may be defended as an idealization: the essential buyer of aspirins, the one of interest to theory, is the man-in-the-street with a headache, not the pharmacologist. So the conditional

[24] One natural regimentation is $(x)(\phi x \rightarrow Pr_k{}^0\gamma x)$, where the quantification is over some suitably broad domain and $Pr_k{}^0$ is read "there is probability k in everyday theory that." But it is not unproblematic, since *inter alia* it makes probability operators have as arguments open formulas.

[25] See Harsanyi, "Games with Incomplete Information Played by Bayesian Players, Part III," *Management Science*, XIV (1968): 486–502. 'Prior' must be understood not as prior to experience but as on the basis of everyday theory alone. The varying belief systems of those about us may be construed as so many diverse "posteriors" produced by the action of private evidence on the common prior of everyday theory.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

probability on the left hand side of (4), C's, may be equated with a conditional probability given in everyday theory:

$$Pr(g_\alpha|f_\alpha) = Pr^0(g_\alpha|f_\alpha).$$

The second assumption is that the everyday generalization about the tendency of Fs to be Gs is, in the first approximation, unaffected by the restriction of the domain of Fs to Fs-offered-in-commodity-trade: everyday theory takes the propensity of Fs to be Gs to be much the same whether or not they are on the market. At least for the more humdrum tradeables, this is surely to be expected, in part because the domain of such Fs spreads in this case over much of the domain implicit in the everyday generalization; most eggs are boxed and traded as eggs. But this amounts to: $Pr^0(g_\alpha|f_\alpha) = Pr^0(G/F)$.

V. LANGUAGE AND TRADE

1. The Semantic Condition. Why should ϕ-offers be made and accepted? In this section, I give an answer to this question and so explain the existence of commodities. The explanation lies in the efficiency of language. I show, on certain background assumptions, that we may expect F to be a commodity rather than to be traded in medieval mode if, and only if, F is denoted by a predicate ϕ which satisfies the following condition, the *semantic condition: ϕ-offers are (i) informative enough and (ii) simple enough.* "Simplicity" will be syntactic. Thus, the semantic condition is the condition that a sort of semantic-syntactic power-weight ratio for ϕ-offers be made sufficiently high by the predicate ϕ.

The notion of the "informativeness" of a ϕ-offer is very slippery. A message may be informative in the sense of inducing true beliefs or in the sense of inducing useful beliefs. I shall adopt a notion that is of the latter sort, the decision-theoretic notion of the *information value*[26] of an "epistemic action."

An epistemic action is one whose only direct consequence for the actor is that it provides her with new information. This information may increase her utility, by enabling her to choose her later actions more advisedly. The information value of an epistemic action is the difference to her utility that would be made by the action's being available (without cost) rather than unavailable. (This improvement may in turn be measured by the maximum sum she would pay to have

[26] The relation between this concept and the information-theoretic one of *amount of information* is explored by K. J. Arrow, "The Value of and Demand for Information," in *Decision and Organization*, C. B. McGuire and R. Radner, eds. (Amsterdam: North Holland, 1972). Dretske's notion of the information content of a signal derives from the latter [F. I. Dretske, *Knowledge and the Flow of Information* (Cambridge: MIT, 1981)].

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the action available.) Here, the actor will be a typical consumer and the epistemic action that of finding out the content of a ϕ-offer. Thus, its information value is the consumer's utility gain from the information that a ϕ-offer contains.[27]

2. *A Model of Two Modes of Trade.* Assume that the *consumer* C is a subject of decision theory whose degrees of desire satisfy

$$Ut(\bar{f}y) = -y,\ Ut(fgy) = p + \alpha - y,\ Ut(f\bar{g}y) = p - \beta - y$$

where a bar denotes negation, y the proposition that C parts with y units of money and f, g as before; and α, $\beta > 0$. Getting no F and paying out y gives utility $-y$; and paying p for an F is preferred or dispreferred to the status quo $\bar{f}0$ according as the F is or is not a G. So G is a positive grader of F.[28]

The market works as follows. All ϕ-offers are at the price p. Every ϕ-offer is accompanied by a second offer, which is just like the ϕ-offer but adds: by paying an "approval premium" of a now, the consumer may subsequently return the delivered object and get the price p back. Thus, accepting the variant offer enables C to inspect an F before definitively buying it. There are N suppliers; for $j = 1, \ldots, N$, the probability is λ that supplier S_j is willing to make a ϕ-offer. Both λ and (N being very large) $(1 - \lambda)^N$ are small.

The courses of action or *strategies* open to C are composed of two sorts of *acts*. There are $N + 1$ *basic acts:* (i) *Not* ("do nothing"); (ii) *Setj* ("settle with S_j") ($j = 1, \ldots, N$). These I shall describe in a moment. There are two *epistemic acts:* (i) *Loc* ("locate"): inform all suppliers that she is interested in receiving ϕ-offers; if any arrive, open one at random. I assume that suppliers answer her advertisement if and only if they are willing ϕ-offerers. (ii) *Ins* ("inspect"): accept the variant of the opened ϕ-offer, and inspect the delivered object. I assume that an inspection settles definitively whether an F is or is not a G. *Loc* costs $c > 0$ and *Ins* costs $d > 0$ (a plus the cost of the physical inspection); the basic acts cost nothing.

C's decision problem has three stages. If at any stage she chooses a basic act, she makes no further choices. At the first stage, she must either do a basic act or *Loc*. At this stage, the act *Setj* consists in informing S_j that she would accept an ϕ-offer from him. If she does *Loc*, then if she gets no replies she must do a basic act; if she does get

[27] Information value is evidently a measure of informativeness relative to the practical context. That of finding out a ϕ-offer generally varies with the consumer's desire for an F, is lower if doing so commits her to accepting it than if all her options stay open, and so on.

[28] I assume that $Ut(fg) = Ut(fg0)$, and similarly for $f\bar{g}$.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

replies she must do either a basic act or *Ins*. At this stage, *Setj* consists in accepting the opened ϕ-offer. *Ins* leads to a third stage, at which she must do a basic act. At this stage, *Setj* means keeping the object she has on approval, and *Not* means returning it.

I assume that C and the suppliers are members of an economy E in whose language ϕ denotes the type F. I write $\pi = Pr^0(G/F)$. I assume C is an F-layperson; then, by (4), C's conditional probability $Pr(g_\alpha|f_\alpha) = \pi$. I assume ϕ-offers are pure: it follows that C attaches probabilities π and $1 - \pi$, respectively, to getting an FG and getting an $F\bar{G}$ were she to accept one. The above-described workings of the market in Fs are trading properties of Fs that I assume to form part of everyday F-theory; all, then, know these workings.

The consumer's decision problem can be represented in a classical *decision tree* (Fig. 1). Spots (D_1, D_{21}, etc.) mark points of decision; a ring (R_1, R_{1j}, etc.) marks a random event, and the numbers next to the branches stemming from it give the probabilities C attaches to the possible outcomes. The tips of the branches of the tree show alternative endings with their utilities.

Two strategies, *COM* and *MED*, are of special interest. *COM* consists of *Loc* followed, if there are no replies, by *Not*, and if there are replies, by *Seto*, where o is the index of the supplier whose offer is opened. In the latter, almost certain, event *COM* results in a ϕ-exchange between C and this supplier, an episode of *commodity trade*. *MED* is *Loc* followed by *Not* if there are no replies and, if there are,

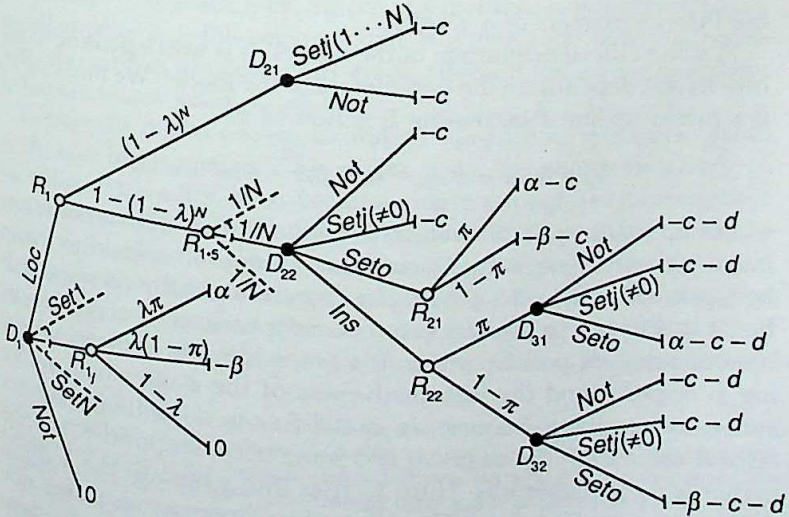


Figure 1

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

by *Ins* followed in turn by *Seto* if inspection reveals an *FG* and by *Not* otherwise. To follow *MED*, therefore, is to engage in *medieval trade*.

This completes the specification of the model. In the next subsection, I establish the semantic condition for commodity trade in the economy it describes: an economy in which consumers are rational agents who face pure ϕ-offers and have, in evaluating them, *F*-laypersons that they are, only the everyday meaning of the predicate ϕ to go on.

3. Proof of the Semantic Condition. The information value for *C* of *Loc* is

$$V(Loc) = U_1 - U_2$$

where U_1, U_2 are the maximum utilities she could attain at D_1 if, respectively, *c* were zero and *Loc* were not available. Now, on the one hand, if *C* does *Loc* she will (with near certainty) learn the contents of a single offer. On the other, if *d* is high enough then evidently she will not inspect, so that the contents of the ϕ-offer are the *only* information that will flow from doing *Loc*. We may therefore take the limiting value of $V(Loc)$ as *d* approaches infinity as a measure of the value to *C* of the information provided by a single offer. Accordingly, I call the quantity

$$V_\phi = lim\ V(Loc) \text{ as } d \to \infty$$

the *informativeness of a ϕ-offer*.

V_ϕ is the crucial magnitude of the model. It is worth pausing to ask how its size depends on the parameters of the model. We find that V_ϕ is a piecewise-linear increasing function of π:

$$V_\phi = 0 \text{ for } \pi \leq \pi_0, = (\alpha + \beta)\pi - \beta \text{ for } \pi \geq \pi_0$$

where $\pi_0 = \beta/(\alpha + \beta)$. The reason is not hard to see. If the consumer had to choose a basic act at D_1, her maximum attainable utility would be (approximately) zero. Getting her hands on ϕ-offers is of value to her if it puts her in a position to choose a basic act—accepting the opened one—of positive utility. If π is too low, the utility of accepting is negative and the informativeness of the ϕ-offer is therefore zero—not negative, because she is still free to do nothing. Above a critical value of π, the utility of accepting becomes positive, and the higher is π the higher it is. Thus, V_ϕ rises with π because the utility of accepting a ϕ-offer does. The latter does because, as π rises, the collateral information about the properties of what is on offer gen-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

erated by its description as an F makes this offering increasingly desirable in virtue of this description.[29]

The *complexity of a ϕ-offer* may be measured in our model by c, the cost of *Loc*. This cost depends on several factors, including the number and geographical distribution of suppliers, and the cost of transmitting a unit of text, such as a character or a phoneme, to different suppliers. But a critical positive determinant is the number $n(\phi)$ of units of text in the predicate ϕ itself, for 'ϕ' must occur in C's text. Since the other determinants of c are fixed in our model, c is a measure in it of $n(\phi)$. And $n(\phi)$ is in its turn the principal positive determinant of the syntactic complexity of a ϕ-offer.

We can show now that acquiring an F by commodity trade is the best option for the consumer if and only if the informativeness of a ϕ-offer is high enough and its complexity low enough; that is, if and only if ϕ satisfies the semantic condition. The proof is in the Appendix.

PROPOSITION. The strategy *COM* is optimal if and only if

$$V_\phi \geq \alpha - \frac{\alpha + \beta}{\beta} d \tag{5}$$

and

$$V_\phi \geq c \tag{6}$$

The proposition has the corollary that there is commodity trade in Fs just if ϕ meets the semantic condition. This follows at once from the rationality of the consumer. Since, *ex hypothesi*, suppliers are happy to sell at price p, it follows, too, that, if ϕ meets the semantic condition, the activity of commodity trade effects a Pareto improvement —the consumer gains, none lose.

Condition (5) puts a lower bound on V_ϕ which varies inversely with d, the cost of inspecting. This makes sense: for commodity trade to be optimal, a ϕ-offer must be informative enough on its own for it not to be worth paying d to make a subsequent inspection; for this to be so, its inherent informativeness must be the greater the lower is d. Condition (6) bounds the complexity of a ϕ-offer; if the information in a ϕ-offer *is* great enough relative to the cost of inspecting, there will be commodity trade if and only if, in addition, its complexity is

[29] Information value is more context-sensitive than a certain intuitive notion, on which a very low value of π makes a ϕ-offer just as informative as a very high one. Information value is low because this is information which would not make the message useful for C to have; but it would if C's grade preferences were reversed. High information value implies a high value for this intuitive notion, but not conversely.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

low enough for the consumer to be able to afford to obtain this information. Fig. 2 illustrates the proposition for $\alpha = 5$ and $\beta = 3$. When $d = ½$, commodity trade is best if and only if the point (V_ϕ, c) lies to the right of the piecewise-linear locus QST; PRT and OT are the corresponding loci when $d = 1$ and when d has any value not less than $\alpha\beta/(\alpha + \beta)$.

4. Relaxations. In my simple model, the semantic condition is necessary and sufficient for F to be a commodity. But simple the model is, and special. In this subsection, therefore, I consider the effect of relaxing some of its main assumptions. I conclude that the model's conclusion is robust to them.

A desired type F may have not two grades but many, even infinitely many. To account well for behavior, it is enough to consider a sufficiently large finite set $\mathcal{G}$. The decision tree now has more branches at the chance nodes R_1, R_{21}, and R_{22}. The central conclusion stands: as in the proposition, there is commodity trade provided that V_ϕ is sufficiently high both relatively to c and absolutely. The way in which the grader information contained in ϕ makes ϕ-offers informative is now more complex, for V_ϕ now depends on $m - 1$ probabilities, if there are m atomic grades. V_ϕ will be high if a thing's being an F means that it is likely to be of many highly prized grades of $\mathcal{G}$.

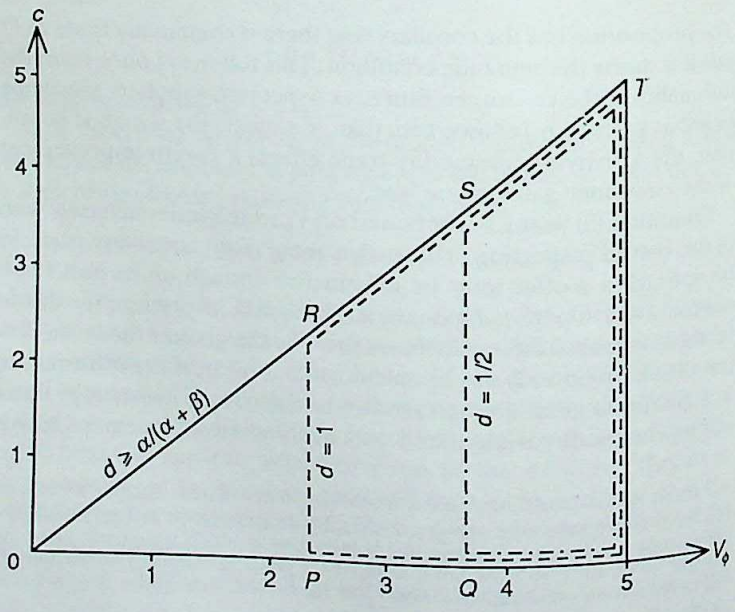


Figure 2

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

There has been interest in the economics of information in trade in a good which has various "qualities." Admitting qualities changes nothing, for qualities of F are just a special kind of grade: namely (if countable), grades FG_1, FG_2, . . . which partition the type F, and are completely ordered in preference.[30]

The process of exchange in our model falls into two phases, a "contact" phase and an "assessment" phase. In the former, a consumer and a supplier make contact, and an offer is made. In the second, the consumer appraises the offer, with or without the help of an inspection. In the model, the consumer initiates things by broadcasting her willingness to buy an F; in reality, the order of moves varies widely. But nothing essential hangs on these variations. If, as is more common, the supplier moves first, say by disseminating offers, he is the proximate bearer of the cost of contact. But, in equilibrium, this cost gets added to the price whether trade is by mere description or after inspection; and the balance of advantage for the consumer between these two modes of acquisition is just as in the model.

The model treats inspection in a rudimentary manner. Letting the consumer's information-gathering decision be governed by a sequential rule rather than be one-shot would clearly change nothing of substance. Of more importance is the assumption that an inspection produces complete certainty. Real inspections produce certainty about some features, such as the vulgarity of a lampshade, but at best sharpen the probabilities for others, such as the remaining lifespan of a gearbox. The more inconclusive is inspection the lower is *its* information value, and the less informative need a ϕ-offer be for commodity trade in Fs to be preferred, for a given cost of inspecting. But commodity trade continues to require that ϕ-offers be informative enough relatively to this cost.

The model shows that, under certain conditions, commodity trade in Fs is best for the consumer when she is confronted with exactly three alternatives relating to F-acquisition—commodity acquisition of one, medieval acquisition of one, or no trade. But, in reality, consumers may learn of there being Fs available under some other true description than as Fs: there may be the option of trading in FGs or $F\bar{G}$s, or in some grosser type of which F is a subtype. The analysis is more complex, but it remains true that, for each type F'

[30] The "quality" literature is about medieval trade; it gives language no role [see e.g., R. Kihlstrom, "A Bayesian Model of Demand for Information about Product Quality," *International Economic Review*, XV, 1 (February 1974): 99–118]. There is no account of how consumers might acquire beliefs about the qualities of Fs other than by inspecting samples of them—in particular, learning the meanings of words describing Fs is not, in this literature, a way in which such beliefs are instilled.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

for sale, the balance of advantage between the two modes is settled by the semantic condition for the predicate ϕ' denoting F'.

The assumption that the consumer trusts the offerer may be thought quite unrealistic, not least by contemporary economists convinced of the ubiquitous rule of Jonas Chuzzlewit's maxim, "Do other men, for they would do you." But it is not groundless. Most commodity trade employs predicates ϕ made up of words ('car', 'tasty') among whose uses such trading uses bulk large, and whose meanings appear stable; and stability of meaning requires a limit to abuse. True, this consideration at best excludes protracted, large-scale misrepresentation—a plague of attempts to palm off wheelless vehicles as cars. But, in addition, misrepresentation is against the law; and suppliers are inhibited from cheating by concern for their reputations, for their ϕ-offers identify them.

These grounds are, however, hardly decisive, and we should ask what difference it would make to drop the assumption of trust. Suppose, then, C thinks it possible she would not receive an F: $Pr(a \,\Box\!\!\rightarrow \bar{f} \mid E) > 0$. Since $Ut(\bar{fp}) < 0$, the curve of V_ϕ as a function of π shifts downward: for some π values V_ϕ no longer makes commodity trade optimal. But the semantic condition is undisturbed: if V_ϕ is high enough *after* being depressed by the dampener of distrust, then as before commodity trade is optimal.

The model's assumptions are of two kinds. Some—the consumer's rationality, the stochastic properties of suppliers' offers, the order of moves—are there for *determinacy*. These are mere "necessities of theory." But others specify conditions that are necessary, like the semantic condition, for *commodity trade*. The latter include the consumer's trust, and her linguistic competence. Imperfect mastery of everyday theory, like distrust, drives V_ϕ down and militates against commodity trade. There is, then, a more general model, in which sufficient degrees of trust and competence join the semantic condition as conditions severally necessary and jointly sufficient for trade by verbal description alone.

VI. CONCLUDING REMARKS

Commodities depend for their existence on words to express the concepts under which fall consumers' desires.

Call the set of the community's verbalized concepts its *conceptual repertoire*. The semantic condition implies that a narrow conceptual repertoire limits the range of possible commodities; an expansion of the former may be a precondition of commodity innovation. A more complete account of commodity trade than I have offered must therefore wait on a theory of the breadth of the conceptual repertoire.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

One aspect of such a theory is capital-theoretic. We may think of the conceptual repertoire as an outfit of tools. As in the case of physical tools, the adoption of an extra concept yields flows of new benefits and costs.

Certain salient sources of such costs and benefits militate for a small conceptual repertoire resistant to invasion by new concepts. If a concept F is to enter the repertoire, people must learn the addendum to everyday theory which treats of Fs—they must "do" a new topic; they must often acquire new recognitional capacities; they must enlarge their vocabulary. All these impose cognitive strain.[31] Once acquired, new concepts must be maintained, expensively, in working order. The larger the existing stock, the greater may be the marginal costs of acquiring and maintaining a new concept.

The possession of concepts brings benefits to the community; commodity concepts allow it to exploit opportunities the world offers it for making Pareto improvements by exchange. There are, moreover, "increasing returns to use." A history of successful use entrenches the concept, and this enhances its utility. Beliefs in generalizations about the grading and trading properties of a type F tend to produce, if the beliefs make the semantic condition hold, commodity trade in Fs. This, where it is net new trade in Fs, increases the flow of new confirming instances of the generalizations. And the more entrenched the beliefs become, the greater are the benefits from, and the more powerful is the spur to, commodity trade.[32]

Once everyday theory is a certain size, the factors I have been discussing tend to eliminate the net benefit from the accretion to it of further commodity concepts. Furthermore, they tend to augment the net benefits, as time goes by, that accrue from established ones. All this leads us to expect a "sparse" and sluggish commodity structure. And indeed, actual commodities are sparsely distributed in the space of Lancaster's "characteristics" (*op. cit.*). Commodity innovation requires in general an *investment* in a concept by the community, having a first negative then positive profile of net returns. The initial costs are likely to make deliberate investments of this kind the preserve of large firms. This makes for the sovereignty of producers, not consumers, in elevating mere collections of objects to the rank of commodities.

My answer to the question—"Why should there be commodi-

[31] See J. S. Bruner, J. J. Goodnow, and G. A. Austin, *A Study of Thinking* (New York: Wiley, 1956).

[32] The simple model of everyday beliefs I have been using cannot show it, but it is clear that the *firmer* is the belief in a tendency of degree π for Fs to be good ones, the greater is the informativeness of a ϕ-offer.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ties?"—implies that, whatever may be our welfare gains from their existence, we have our language to thank for these gains. Language stands side by side with morality and law as a social artifact that plays an essential part in our reaping of the benefits of exchange. My answer thus opposes the myth of liberalism, the doctrine that economic welfare is the product of autonomous individual choices *in vacuo*. Ordinary language and ordinary thought have been the victims of neglect in economic science in general. Their potential for the explanation of the most economic of phenomena may, however, be great. It is, I hope, exemplified by the present semantic theory of commodities.

MICHAEL BACHARACH

Christ Church, Oxford

APPENDIX

Proof of Proposition. We solve the problem represented in Fig. 1 by the standard method of backward recursion based on Bellman's principle of optimality. At D_{31}, *Seto* is best and has (in obvious notation) value $Ut(Seto|D_3) = \alpha - c - d$; at D_{32}, *Not* is best, with value $-c - d$. Thus,

$$Ut(Ins|D_{22}) = \pi(\alpha - c - d) + (1 - \pi)(-c - d) = \pi\alpha - c - d$$

For the other options at D_{22} we have

$$Ut(Seto|D_{22}) = \pi\alpha - (1 - \pi)\beta - c$$

$$Ut(Setj|D_{22}) = -c \quad (j \neq o)$$

$$Ut(Not|D_{22}) = -c$$

Thus, the value of reaching D_{22} (i.e., of opening an offer S_o) is

$$max\ \{-c,\ \pi\alpha - (1 - \pi)\beta - c,\ \pi\alpha - c - d\}$$

Since by symmetry this is also the value of opening an offer from any S_j, and the probability of opening an offer is $1 - (1 - \lambda)^N \simeq 1$, the values of the options at D_1 are

$$(7) \qquad Ut(Loc|D_1) \simeq max\ \{-c,\ \pi\alpha - (1 - \pi)\beta - c,\ \pi\alpha - c - d\}$$

$$Ut(Setj|D_1) \simeq 0$$

$$Ut(Not|D_1) = 0$$

Thus, *COM* (do *Loc* at D_1 intending *Seto* if D_{22} is reached) is best if and only if

$$(8) \qquad \pi \geq (\beta + c)/(\alpha + \beta)$$

and

$$(9) \qquad \pi \geq 1 - (d/\beta)$$

It is easily shown that (8) and (9) are equivalent to (5) and (6).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

COMMENTS AND CRITICISM

MISUNDERSTANDINGS OF EPISTEMIC TIT FOR TAT: REPLY TO JOHN WOODS

In John Woods's[1] comment on my paper,[2] he takes me to task for attempting to illuminate the trust and cooperation necessary for epistemic collectivism by modeling it analogically as an iterated prisoner's dilemma (PD): first, by the production of purported counter examples and, second, by claiming that the model is not mathematically formalizable. Woods's comments are worthy of reply and I shall address them, welcoming this opportunity to clarify some of the essential points.

I. THE COUNTER EXAMPLES

I-A. Technical Details. Woods reminds us that decisional nodes of PD-play are opaque, in the sense that the exchange of communication at any given node is a function of past actions or speech acts so that the contributions of any two individuals is at best an unordered pair, whereas true communication nodes should be modeled as ordered pairs of responses that do not simple co-occur at that node but are partial functions of the elements occurring at that very node (326–7). Although it may very well be useful to think of the exchange of information as Woods would have it, it is not correct to suppose that the application of TIT FOR TAT to the knowledge game implies that the exchange of information at any given node need take account of anything but past exchanges. In casual or formal communication, a person can only respond to what has gone before; even high-speed computerized transfer of information requires alternating, not simultaneous, exchange. I believe that it is only for one's own artificial convenience that we can speak of simultaneous exchange at a given node. Dialogue cannot be simultaneous; it alternates.

I-B. A Lie for a Lie. The first real proposed counter example takes the form of an imaginary case involving two senior officials attached to different departments within the same government. They have agreed to exchange sensitive information by blind drops at a pre-arranged location (327–8). Woods recognizes that, if the exchange is strictly one-shot, "then the classical PD-maneuver would indeed secure a purchase" (328). If the exchange is iterated, then a strategist

[1] "The Maladroitness of Epistemic TIT FOR TAT," this JOURNAL, LXXXVI, 6 (June 1989): 324–331.
[2] "Epistemic TIT FOR TAT," this JOURNAL, LXXXIV, 8 (August 1987): 363–376.

0022-362X/90/8707/369–374 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

employing TIT FOR TAT may win against a TESTER strategist, or against another TIT FOR TAT player, but certainly not against an ALL D strategist who defects continually by leaving only misinformation at his drop. TIT FOR TAT will reply to ALL D's continual defection by defecting all the time in return.

Woods says that, in such a case, the rational course is not to continue defecting *ad infinitum,* but to quit playing against ALL D and find another partner. With this, I heartily agree. Exactly at what point should the TIT FOR TAT strategist decide that her opponent is going to continue defecting and that the agreement should be terminated? This, the strategy itself cannot say; it is up to the strategist to weigh results and ponder more effective strategies. I shall return to this point later: my reply to the next counter example should clarify it.

I-C. The Unscrupulous Researcher. Another counter example involves the relationship between an individual and a scientific journal; Woods thinks that such a relationship is not well-described in PD-terms, whether we look at a one-shot exchange or an iterated interaction. In the one-shot case, Woods says that the "best course for an author in a one-shot submission to a scientific journal is to send off a cheating piece of work, and the best course for the journal is to turn it down sight unseen" (329). I agree that in a noniterated PD-environment, the author's best bet is, indeed, to defect in such a manner; but a journal's defection does not consist in rejecting the piece sight unseen. Woods errs in improperly construing cooperation and defection on the part of the journal.

When two players (an author and a journal) play the knowledge game within the scientific framework, each cooperates by doing the job well and defects by not doing it well. For the author, doing his job well (cooperation) consists in submitting only correct results that have not been plagiarized, to the best of his knowledge; not doing his job well (defection) consists in cheating. On this, Woods and I are in good agreement. For the journal, cooperation consists in correctly reviewing submitted papers, publishing good papers as space permits, rejecting bad ones, refraining from nepotism, and, in general, being as objective as possible; defection, on the other hand, consists in haphazard reviewing, nepotism, favoritism, nonobjectivity, refusing to publish good papers for bad reasons, or publishing shoddy material. A journal has not defected just because it refuses to publish a paper; on the contrary, the journal is doing its job—is cooperating—when it rejects a bad paper. The journal defects when it refuses to publish a good paper, not when it simply refuses to publish.

In the iterated case of an author playing with (against?) a journal,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Woods again incorrectly takes the journal's cooperation to be acceptance and its defection to be rejection of submitted papers, thus concluding that the best journals are not TIT FOR TAT players at all, but rather ALL D players because they reject better than ninety percent of submitted papers. I would maintain, on the contrary, that the best journals play a form of TIT FOR TAT. Individual researchers do so also, for they become wary of researchers whose results are not readily replicable. Also, rejecting better than ninety percent of received papers is not, contrary to what Woods suggests (329), an ALL D strategy, but rather something that should be called MOST D.

In the knowledge game, reward is recognition, punishment is the withholding or the withdrawal of recognition. This is true for individual researchers, and also for collective entities like a journal or a research institute. An individual's fame is based largely on his record of publication (his cooperation in the game), and his loss of reputation is due to his bad record (defection). A journal's reputation (reward) is enhanced by being tough but fair in its treatment of submitted articles; it is destroyed (punishment) by bad editorial decisions. The following payoff matrix summarizes this interpretation.

		Player B (scientific journal)	
		Cooperates (publishes good papers, rejects bad ones, relies on reputable peer review, etc.)	***Defects*** (publishes bad papers, rejects good ones, practices nepotism or favoritism, etc.)
Player A (individual researcher)	***Cooperates*** (submits good papers, does not fabricate data, etc.)	(gain or maintain reputation, gain or maintain reputation)	(gain or maintain reputation, lose reputation)
	Defects (submits bad papers, fudges data, practices plagiarism, etc.)	(lose reputation, gain or maintain reputation)	(lose reputation, lose reputation)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

In order to understand better why both of the foregoing counter examples that Woods has proposed do not count against my own proposal, it is important to realize that refusing to play well is defection, whereas outright refusal to play is neither cooperation nor defection, but simply stepping out of the game altogether.

As a strategy, TIT FOR TAT enjoins its employer to follow a certain course of action; followed blindly it may win, but it may also lose, and lose heavily, against some opponents. Decisions whether to continue to use the TIT FOR TAT strategy belong to the strategist, not to the strategy itself (it has no provisions for such a decision—followed precisely, it will simply continue cooperating or defecting according to the other player's previous moves).

The strategist's decision to employ or cease to employ TIT FOR TAT or any other strategy is not modeled at all in my original paper. Rather, I attempted to show how this particular strategy could be illuminating for epistemic collectivism where no moral sense of trust intrudes; even without morality, I tried to indicate, cooperative behavior is possible and explainable by reference to such a strategy as TIT FOR TAT. This cooperative behavior could be elicited simply be reference to reward-seeking and punishment-aversion behaviors in individuals as well as in collective entities.

I-D. The Check Bouncer. Woods produces a last counter example that has nothing to do with knowledge at all, but rather with commerce. This counter example seeks to show, I take it, that PD-constraints do not apply to the real world at all, whether in commercial or in scientific domains. Again, I believe Woods has misconstrued how to interpret cooperation and defection in this context, just as he did in the previous two.

In a one-shot transaction, it is indeed in the interest of the individual buyer to try to pay for a purchase with a bad check; and it would also be in the interest of the store to furnish shoddy goods instead of quality items. Shady deals are particularly prone to this sort of double defection: con artists are past masters at it, as are, if we believe novelists like John LeCarré, shifty spies.

In an iterated transaction, the store obviously wishes to do business again with each customer, and hopes its reputation will attract new ones; so it is not in its interest to defect by selling shoddy goods, or by substituting worthless items for good ones in sealed boxes [which is one mode of defection suggested by Woods (329)]. In such a situation, the crooked individual can indeed try to get away with passing bad checks. The store could continue to cooperate with the crook normally, and simply take the loss. More probably, the store may withhold the item until the crook's check has cleared. The store

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

may also continue to do business with the crook, but only on a strictly cash basis. Or, finally, it can simply refuse to play the game with the crook.

It is not in the long-term interest of the store to defect with any customer, for it will lose its reputation with that customer, and with others through word of mouth. It will thus continue to supply quality goods to everyone. Since many good customers seem to dislike cash transactions, the store welcomes checks. It may choose to require only perfunctory identification before taking an individual's check, accepting a bit of a loss due to bad checks rather than inconvenience every customer with elaborate on-the-spot credit checks. In such a case, the store cooperates with each customer, most customers in turn cooperate by writing good checks, and only a small, economically acceptable, minority of customers attempts cheating. The wise store manager will identify habitual offenders, and stop accepting checks for payment of goods; if she sells for cash only, then the store is still cooperating; if she decides to no longer do business with the habitual defector, then the store has decided—not to defect!—but to stop playing altogether.

II. FORMALIZABILITY

It is true that my proposal is not mathematically formalized. But Woods himself describes its spirit quite well:

> Why not understand Blais's as a commitment to what we might call pseudo TIT FOR TAT, a strategy stripped of all its PD-trimmings and defined for any cooperative situation in which cooperation and defection can realistically be specified, and in which only the following features of (real) TIT FOR TAT are mimicked: pseudo TIT FOR TAT cooperates first, defects only to repay a defection, repays a given defection always but only once, and cooperates otherwise? So interpreted, the Blais thesis would now be that pseudo TIT FOR TAT represents and illuminates norms of epistemic cooperation (330).

But Woods cannot accept this, because, he claims, Robert Axelrod's[3] studies have produced results that tell a "dramatic story [that] is told in the mathematics of the PD-ecologies and is inseparable from, inheres in, the mathematics" (330). But what, I wonder, does Woods make of the fact that better than half of Axelrod's seminal book is concerned with illustrations and applications of the virtues of TIT FOR TAT (it is a *nice, provocable, forgiving, clear,* and *robust* strategy that is all the more efficient if the *shadow of the future* is long) in fields as diverse as biology (ch. 5), legislative assemblies and commerce (ch. 6), and even trench warfare (ch. 4), to name only these examples? It is true that Axelrod does not mention the

[3] *The Evolution of Cooperation* (New York: Basic, 1984).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

world of scientific knowledge as a possible sphere of application. But that is precisely why I chose to explore an application of Axelrod's insights to what I thought to be an interesting new domain.

Woods requires that any application of Axelrod's theory to non-PD-environments meet formal mathematical criteria, else the mathematics will lose its purchase in the absence of "an appropriate representation theorem giving representation functions that take PD-ecologies into non-PD-ecologies and yet preserve the mathematics" (331). I have found Axelrod's own analogical applications of his formal results to real-life situations to be extremely interesting, although no formal mapping such as Woods requires of me is provided. All I can say is that Woods is obviously free to refuse my extension to the epistemic domain, if he so chooses. It would be interesting to know if he similarly castigates Axelrod for his lack of formal mapping.

III. CONCLUSION

Although I did not try (because I do not believe it to be possible) to formalize my model of epistemic TIT FOR TAT in the original "Comment and Criticism," I did supply the pertinent epistemic interpretations of the key terms of the formal TIT FOR TAT strategy (cooperation, defection, reward, etc.). And, I have tried to clarify this interpretation in this rejoinder to Woods's critique. What is most interesting, I think, is that moral trust is not a necessary foundation for the reliability of the accumulating knowledge afforded by the methods of cooperative scientific investigation, that, as Axelrod states, "[T]he foundation of cooperation is not really trust, but the durability of the relationship" (*ibid.*, p. 182).

MICHEL J. BLAIS

Université de Sherbrooke

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

BOOK REVIEWS

A Treatise of Social Justice, Vol. I: Theories of Justice. BRIAN BARRY. Berkeley: University of California Press, 1989. xv, 428 p. Cloth $45.00.

In this trilogy, Barry is concerned with justice, which, following Rawls, he conceives as a property of basic social institutions (146). He distinguishes two main functions that can be attributed to justice so understood: either "the function of justice is to provide a rational basis for agreement among people who . . . seek to take due account of the interests of all" or it "is to enable egoists to get along better with one another" (272). These two views of the function of justice are correlated with two substantive approaches: justice as impartiality and justice as mutual advantage.

In discussing these two approaches, Barry further confines himself to, what he calls, constructivist theories of justice. These are theories that select a criterion of justice, or just practices, in an indirect (259), broadly contractualist (269) way, as "what the parties would themselves agree on in some hypothetical situation" (269). Barry contrasts this with "intuitionist" theories favoring "the direct route, which would have us simply say what we think justice requires" (259).[1]

Constructivist theories of justice as impartiality are dubbed original-position theories. Barry classifies these into four types, distinguished on the basis of two questions (294, 320f): Do the parties know their (clients') personal characteristics? And are they motivated by self-interest or by a desire to reach agreement on reasonable terms? In the second volume of his trilogy, Barry intends to develop an original-position theory that conceives the parties—though "they do not have to be completely detached from their own interests" (352)—as motivated by a desire to reach agreement on reasonable

[1] Barry also maintains, however, that constructivist theories are intuitionist by relying on intuitions (271–5). He might have been clearer here that "intuitionism" may name two distinct doctrines. One holds that value conflicts can be resolved only intuitively through case-by-case judgment, not algorithmically through priority rules or assignment of weights. The other holds that our moral reflections must be based on certain intuitions as fixed points (to be understood as glimpses of the moral facts, as in G. E. Moore, or as considered judgments, as in Rawls). Barry also wavers on whether the latter doctrine allows appeal to intuitions on all levels of generality (272f) or only to "intuitions about well-described particular cases" [which, falsely, he takes to be Rawls's "official theory" (281)].

0022-362X/90/8707/375–384
© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

terms.[2] This original position is to be a debate rather than a game (371), though (in keeping with the definition of "constructivism") a fictional debate among "agents of construction" (371). These agents "do not operate behind a veil of ignorance" (371), though such a veil will be invoked occasionally for dramatic emphasis (333, 345, 371). Does this mean that reference to a veil will occasionally be made by the debating parties, or that Barry will occasionally spread a veil over their debate? Will Barry's theory resemble that of Jürgen Habermas or of Bruce Ackerman, who also reject the use of a veil? While awaiting the sequel, we can discuss the main goal of the present volume, which is to analyze and discredit some competing constructivist theories of justice.

Barry takes seriously competing original-position theories of only one type—those which conceive the parties as self-interested and ignorant of their (clients') personal characteristics. We are told, however, that this combination is bound to constitute a "mismatch between the fundamental intuition about the separateness of persons and the machinery of rational choice under uncertainty" (336); "impartial morality does not emerge from self-interested choices, however we rig the rest of the specification of the original position" (334). This objection misses the dominant *moral* rationale behind the assumption of self-interest (or, more accurately, of mutual disinterest). This rationale is clearest when the parties are conceived as *representatives of* prospective citizens: the parties' single-minded concern for their clients' interests (which need not be selfish ones) reflects the idea that, to be just, social institutions ought to accommodate and protect the interests of all and only those who are to live under them.

Barry further worries that such theories pack so much into their premises (their original position) that these premises do not have enough independent plausibility to increase our confidence in the favored criterion of justice. In leveling this charge against John Harsanyi (334f), Barry fails to point to particular premises. When he levels it against Rawls (336–340), he claims that, by stipulating "that the people for whom principles are to be chosen have an overriding desire to be just" (337), Rawls makes his entire construction irrelevant, for "if you do not in fact have the higher-order interests that Rawls attributes to you, then a theory premised on the assumption

[2] This desire is borrowed from T. M. Scanlon's "Contractualism and Utilitarianism," in A. K. Sen and B. Williams, eds. *Utilitarianism and Beyond* (New York: Cambridge, 1982), pp. 103–128.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that you do have them cannot be of any concern to you" (338). But what Rawls stipulates is not that those whom the parties represent *have* a sense of justice, but that they have an *interest* in having it, i.e., an interest in institutions that will engender it. This stipulation is plausible because such institutions are good for citizens. We all have strong reasons to value a sense of justice in one another.[3]

I

Constructivist theories of justice as mutual advantage, which seek to appeal only to rational prudence and self-interest, are dubbed two-stage theories. The mutual advantage approach has received a great impetus from modern game theory, and Barry spends the entire first third of his book discussing two-person bargaining games. I found this discussion competent and helpful in locating the differences among John Nash, R. B. Braithwaite, John Lucas, David Gauthier, and others. It also provides (though Barry no more than hints at this) one reason for being skeptical about two-stage theories: none of the several supposedly unique general solutions to bargaining problems now in the field shows much promise of trouncing all its competitors. Likely to prevail is therefore the claim, made already in 1944 by John von Neumann and Oskar Morgenstern (12f), that there is no unique solution. This would presumably entail that there is no uniquely acceptable two-stage theory of justice either.

But two-stage theories face a deeper challenge as well, one that tends to be obscured by the exciting parallels with game theory. Barry assumes that, being constructivist, such a theory will seek to show what agreement it would be mutually advantageous to make in some particular *hypothetical* bargaining setting (266, 269). It will do this (hence the label) in two stages: it will first establish the noncooperative baseline (the payoffs resulting from how parties would behave before any agreement), and then proceed from there to an agreement on the Pareto frontier. Yet this is not enough, for such a theory must also motivate the hypothetical bargaining setting itself, must show that actual parties here and now have reason to care

[3] For the details of this response, see for example my *Realizing Rawls* (Ithaca: Cornell, 1989), pp. 99–103. I also review there (pp. 111f) some reasons for doubting the old charge, repeated by Barry (215, 340), that, barring strong new stipulations, Rawlsian parties would adopt an averaging rather than a maximin criterion. Briefly, a maximin criterion is easier to apply, better supports a shared sense of equal citizenship, engenders weaker strains of commitment, and for these reasons is more likely to foster stability. Hence, it may well be chosen over an averaging criterion even by parties concerned to maximize their expected payoffs (as a resourcist criterion may well be chosen over a welfarist one even by parties ultimately concerned with welfare).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

about what bargain would be struck in the idealized setting.[4] That some hypothetical contractors would be fools not to set up market institutions, for example, does not show that the same is true of our Mssrs. Ceausescu or Kim Il Sung. Seeing that our world is quite different from any textbook bargaining setting, they, and anyone, should surely ask: What does that hypothetical bargaining have to do with me?

This question—though a staple in discussions of original-position theories—is too often neglected in discussions of the mutual-advantage approach. Barry is no exception. At the outset of his discussion of original-position theories he writes: "Unless we have been given some good reason for believing that what would be chosen in a certain kind of situation will constitute a principle of justice, it is a waste of time even to inquire about what might be chosen in it" (321). But his discussion of mutual-advantage theories starts out as follows: "The first thing that has to be settled is the nature of the nonagreement baseline: Should it depend on the results of strategic play for relative advantage, or should it reflect the best the parties could do for themselves independently?" (293f). Barry fails to realize that two-stage theories (as he calls them) must have *three* stages: (A) an argument for some particular hypothetical bargaining setting, (B) an argument for a precooperative baseline appropriate to this setting, and (C) an argument for the bargain that, given this baseline, would be struck in that setting.[5]

It seems difficult, within the mutual-advantage approach, to motivate the favored hypothetical bargaining setting while appealing only to the diverse interests and rational prudence of fully situated parties.[6] Failing this, one might invoke moral considerations—though this would undermine the key rationale behind a theory of justice as

[4] This challenge is quite different from the question, which Barry considers (367), why a merely hypothetical *agreement* should be binding on those for whom it would have been rational to make.

[5] While Barry often flatly overlooks stage (A), he occasionally (302f, 327, 368) seems to conflate it with stage (B), e.g., by using the word 'baseline' with a double meaning.

[6] Only Hobbes seems to me to have a serious response in this vein: my hypothetical bargaining setting incorporates assumptions about the human environment and psychology which are realistic, so that the two main outcomes for my fictional bargainers are also the two main outcomes in the real world. Of course, human persons differ widely in their endowments. The state of nature, however, is a disastrous outcome for all, whatever their human endowments; and the alternative outcome can be achieved only if all persons subject themselves equally to a common power. Hence, it would be irrational, even for the best-endowed actual persons, to try to exact more favorable terms from their sovereign or compatriots.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

mutual advantage. Instead, one might also radically confine one's ambitions, seeking merely to show how, *given* some particular *actual* bargaining setting, a just agreement is to be derived for that setting.[7] But even merely offering a general recipe for such derivations would not fully avoid the problem, because the recipe would still be based on the assumption that all parties in such actual settings are fully rational maximizers. But this assumption is not true, nor is it necessarily the case that each party has sufficient reason to want all to become such maximizers.

Barry wants to classify mutual-advantage theories into four types (294f). In theories of type 1, both the precooperative baseline and the division of cooperative gains are determined by how rationally self-interested parties would bargain with one another. The other three types differ from type 1 by making a stipulation that is not based on what would emerge among rational self-interested parties, namely, the stipulation that the parties make no threats, agree to equal gains (in terms of utility), or both. Barry has little trouble showing that, in making such stipulations, theories of the latter three types fall outside the mutual-advantage approach. But the point of this achievement is unclear. With only one type remaining, Barry's classification becomes idle. And the theories he expels from the mutual-advantage approach never find a new home. (Although clearly constructivist in his sense, they do not appear among the original-position theories.)

Moreover, an important recent theory is expelled unjustly. Barry acknowledges (296) that Gauthier's advocacy of equal gains (strictly: minimax relative concession) does not place Gauthier outside the mutual-advantage approach, because this advocacy is based not on the claim that equal gains is "ethically attractive," but on the claim that it is the rational bargain to strike. Yet Barry fails to acknowledge (296) the parallel point about Gauthier's exclusion of threats (his "Lockean" proviso), which likewise is not based on ethical attractiveness, but on the claim that it is not advantageous, among rational parties, to honor and make threats.[8] Even if Gauthier's argument for

[7] Such a genuinely two-stage theory would make justice extremely context-dependent (more so even than utilitarianism does). No practice would be just simpliciter; and almost any practice advantageous to a few would be just in some context, e.g., when it is disadvantageous for the others to rebel. (What international order is just, for example, would vary dramatically with shifts in the international distribution of power.)

[8] See Gauthier, *Morals by Agreement* (New York: Oxford, 1986), esp. p. 195. Let me note that Gauthier's proviso is extremely weak. It imposes upon the precooper-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

this latter point is unconvincing, this cannot alter the fact that his theory has the type-1 structure and thus properly instantiates the mutual-advantage approach.

II

Thus far, I have concentrated on the final, third part of Barry's book which, for his purpose in this trilogy, is clearly the most important. I have briefly mentioned the first part, a state-of-the-art discussion of two-person bargaining theory. Let me now say a little about the second part, which interprets both Hume and Rawls as committed to two conflicting theories of justice, one within the impartiality and one within the mutual-advantage approach (152). This part is, I think, the weakest of the book because, mainly exegetical in nature, it is just too deeply wrong about the works it is interpreting. I shall briefly address some of the more important points.

Hume's main object is to "introduce the experimental method of reasoning into moral subjects"; he wants to be anatomist rather than painter, i.e., *explain* morality rather than preach it.[9] Not being clear about this distinction, Barry often takes Hume to be involved in Barry's own project of justifying (and appraising) practices, rather than in that of explaining existing practices and human attitudes toward them. Of course, these two projects are not entirely independent. Motives appealed to in justifications may also figure in explanations. The inverse, however, is not generally true, as the case of unconscious motives attests. It is not surprising, then, that Barry sees flaws in Hume's account (which he explains as conservative bias) precisely at points where Hume's explanation invokes subtle psychological mechanisms that could not possibly play a justificatory role.

Although an explanation, Hume's account is still exposed to a possible conflict of the kind Barry envisions, for it is on the basis of self-interest (combined with sympathy and reflection) that Hume seeks to explain the prevalent commitment to a conception of justice as impartiality. "The question is: why should we want to bring our own sympathetic judgments into line with those of everybody else?

---

ative baseline the constraint that *A* not make *B*'s situation worse than it would be in *A*'s absence except insofar as is necessary to prevent *A*'s situation from becoming worse than it would be in *B*'s absence (Gauthier, pp. 203–5). Thus, in the case of Jonathan and Joanna stranded on an island (Gauthier, p. 207), he may prevent her from eating any naturally available food, because her eating would reduce his choice of foods and thus make him worse off than he would be in her absence. By contrast, neither Locke nor even Nozick would allow Jonathan to starve Joanna in this way.

[9] Hume: *A Treatise of Human Nature* (New York: Oxford, 1973), subtitle and pp. 620f.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

. . . In the end, I think that Hume was forced to abandon his official theory and allow that the desire to behave in a way that can be justified in impersonal terms must be admitted as an irreducible motive" (166f). But Hume has a response to this version of Barry's conflict charge [though Barry denies that this response is in Hume (165f)]: it is because human beings are "sensible of" the mutual advantageousness of uniform dispositions of praising and blaming that they tend to equilibrate toward a common moral vantage point —perhaps unconsciously, like two rowers who may fall in with each other without being fully aware that this is the most effective method of rowing (cf. *Treatise*, p. 490).

> 'Tis impossible we cou'd ever converse together on any reasonable terms, were each of us to consider characters and persons, only as they appear from his peculiar point of view. In order, therefore, to prevent those continual *contradictions*, and arrive at a more *stable* judgment of things, we fix on some *steady* and *general* points of view; and always, in our thoughts, place ourselves in them, whatever may be our present situation (*Treatise*, p. 581f).[10]

Barry's conflation of explanation and justification is most damaging in his discussion of the circumstances of justice. These are very general facts about human beings and their environment which, according to Hume and Rawls, must obtain if the problem of justice (parties making conflicting claims on the design of common practices) is to arise at all. Barry thinks, however, that these circumstances also constrain how the problem, once it does arise, should be resolved. Thus, where Hume and Rawls assume that common practices will exist only in an environment in which mutually advantageous cooperation is possible, Barry takes them also to demand that such practices be mutually advantageous:

> If the reason for adhering to the solution is self-interest, as the doctrine of the circumstances of justice implies, the solution must first give the parties what they could get at the nonagreement point; but it must then divide the gains that are to be made by moving away from the nonagreement point in a way that leaves them equally satisfied, in a sense of 'equally satisfied' that takes account of bargaining strength (186f, cp. 249).

The mistake here lies in the supposed implication. The *causal* claim that cooperative practices will exist only in environments in which

[10] Cf. Hume, *Enquiries Concerning Human Understanding and Concerning the Principles of Morals* (New York: Oxford, 1972), p. 229.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

some such practices can yield a net benefit does not imply that the participants' *reason* for compliance with cooperative practices is self-interest. (Compare: the claim that monotheistic beliefs will exist only in large societies asserts a necessary *causal* condition. It does not imply that people's *reason* for holding monotheistic beliefs is that their society is a large one.[11]) Thus, Barry fails to show that their view on the circumstances of justice commits Hume and Rawls to the central normative demand of mutual-advantage theories of justice (which then allegedly conflicts with the commitment to impartiality ascribed to those same authors).

The mistake is further compounded. For Hume and Rawls, whether the circumstances of justice obtain depends on our psychology, environment, and (to some extent) on the general level of technological advance. It does not depend on how a society is organized once these factors are given. Barry, however, takes the circumstances of justice to be predicated of full-fledged social systems, and thus concludes that these circumstances may not obtain *in some societies* because, for example, one group is capable of enslaving others by virtue of its superior technology. It is, Barry reasons, "an immediate consequence of Hume's position" that such enslavement would not be unjust, because one of the circumstances of justice, rough equality of powers, is absent (162). But in the passage from which Barry derives his "immediate consequence," Hume does not make a *normative* point about *any*, but a *factual* point about *natural*, inequality of powers:

> Were there a species of creatures intermingled with men, which . . . were possessed of such inferior strength, both of body and mind, that they were incapable of all resistance, and could never . . . make us feel the effects of their resentment; . . . the restraints of justice and property . . . would never have place in so unequal a confederacy (*Enquiries*, pp. 190f).[12]

Barry misunderstands Rawls analogously. The circumstances of justice would not obtain, Rawls holds, if there were no feasible mode

[11] The mistake, obvious here, is often obscured by Barry's use of ambiguous language, such as 'because', which may indicate a cause or a reason, or 'the requirement of mutual advantage' (249), which may allude either to the factual precondition that mutually advantageous practices must be possible or to the demand that cooperative practices ought to be mutually advantageous.

[12] Yet Barry is right to ask (190) how Rawls can speak of justice between generations, given that future persons can never make us feel the effects of their resentment. This problem, to which Rawls seems indeed to have given no adequate response, has however been raised before—e.g., by D. C. Hubin, "Justice and Future Generations," *Philosophy and Public Affairs*, IV (1976): 70–83, pp. 79f.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of cooperation that achieves a sharable net benefit relative to non-cooperation. But Barry reads Rawls as committed to the claim that the circumstances of justice do not obtain in a society if a transition to just institutions would not be advantageous to those presently rich and powerful (326f, 330). Thus, Barry concludes, Rawls must presuppose "that the principles of international morality that he proposes would really be found mutually advantageous by all countries irrespective of power" (246)[13] and is committed to the demand "that principles of justice should be capable of being presented as mutually advantageous in the real world. This is precisely the significance of the doctrine of the circumstances of justice" (195).

The idea that his endorsement of Hume's circumstances of justice commits Rawls to a mutual-advantage theory of justice is adventurous enough. Yet even more fantastic is Barry's notion that Rawls had initially intended the circumstances of justice as a representation of impartiality. In his earlier article "Justice as Fairness,"[14] "Rawls says that the 'typical circumstances of justice' represent 'the constraints of having a morality' " (325). Here even the quotations are invented. What Rawls actually says is:

> The character and respective situations of the parties reflect the typical circumstances in which questions of justice arise. The procedure whereby principles are proposed and acknowledged represents constraints, analogous to those of having a morality, whereby rational and mutually self-interested persons are brought to act reasonably (*ibid.*, p. 172).

Thus, Barry is wrong to claim that here "fairness is defined by the presence of the circumstances of justice" (325) and "the circumstances of justice are put into the original position" (329). Fairness is represented by the "procedure whereby principles are proposed"; and "the typical circumstances in which questions of justice arise" are those of everyday life. The original position, in any case, does not yet appear in "Justice as Fairness."

The misreadings of Hume and Rawls reveal a pervasive feature of Barry's book: though composed in the assuring style of an elder statesman, it is in fact written with little care and riddled with mistakes. A few further examples must suffice. (1) Although about half the book (and much of Barry's earlier work) is on Rawls, Barry cites only three of the thirteen or so papers Rawls has published since *A*

[13] An earlier version of this claim by Barry is discussed in my *Realizing Rawls*, pp. 105–146, 263–5.

[14] *Philosophical Review*, LXVII (April 1958): 164–198.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Theory of Justice. He also fails to consult the revised text from which translations have been made since 1975. It is then not surprising that he repeats misunderstandings of Rawls's theory that have long been cleared up.[15] (2) Barry finds highly suggestive that the idea of conceiving of the original position as an instance of pure procedural justice does not yet occur in *A Theory of Justice* (267f, 278). One glance at the index of Rawls's book ("Procedural justice, pure . . . and original position") would have given him two passages where this idea does occur. (3) Barry asserts that "the difference principle picks out the most egalitarian of all the Pareto-optimal arrangements satisfying the requirement that everyone should gain from inequality" (227). Shortly thereafter (230), he himself describes a case (without chain connection) that is a counterexample to this assertion. (4) The book would have benefitted from proper proofreading, quote checking, and a more reliable index.

Having focused this review on the weaknesses of Barry's book (as I think he would have wanted a reviewer to do), I have probably made it appear worse than it is. Still, it really is not a strong book—not in a class with Barry's earlier *Political Argument*.

THOMAS W. POGGE

Columbia University

[15] Another example of this is Barry's comment on Rawls's assumption that "the parties take no interest in one another's interests" (*A Theory of Justice*, p. 127): "It is important to realize that the 'parties' being referred to here are not the people in the original position . . . [but] people in real life" (181). The one passage that suggests this reading (p. 128) is corrected in the translations. And many other passages militate against Barry's reading: Rawls speaks of "the postulate of mutual disinterest in the original position" (p. 129); and he later spends an entire section (§79) explaining that mutual disinterest must not be understood "as describing a certain kind of social order . . . that is actually realized" (p. 520), that his ideal is not that of a "private society" (p. 521f) but that of a "social union of social unions" (pp. 527–9).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVII, NO. 8, AUGUST 1990

IDENTITY STATEMENTS AND THE NECESSARY A POSTERIORI

CAN essentialism be proved? Can it be shown by valid argument from uncontroversial premises that certain things, or certain kinds of things, have some of their properties necessarily? Or must we accept that, in the end, essentialist claims need to rely on appeals to our intuitions via Twin-earth examples and the like, appeals to "what we would say" in particular hypothetical situations?

There is a form of argument for a certain kind of essentialist conclusion which seems not to depend upon any appeal to intuition. Identity statements involving natural-kind terms are frequently adverted to in the philosophical literature as examples of the necessary a posteriori, and it can appear, given the nature of the arguments that are put forward in their favor, that the essentialist is on very strong ground with respect to these claims. It is not merely that they are apt to strike one as plausible in the light of modal intuitions or philosophical arguments; they appear to be provable from fairly trivial logical results together with uncontroversial empirical truths. The anti-essentialist is left challenging, at the risk of absurdity, the joint verdict of science and logic. It is with this kind of essentialist claim which I shall be concerned in this paper; I shall be attempting to counter the suggestion, made particularly convincing by Saul Kripke, that essentialist claims of this variety are provable.[1]

[1] My conclusion will thus be broadly similar to that reached by Nathan Salmon in "How Not To Derive Essentialism from the Theory of Reference," this JOURNAL, LXXVI, 12 (December 1979):703–725 and in *Reference and Essence* (New York: Blackwell, 1982). In the first of these, Salmon argues that essentialism cannot be derived from the theory of reference; but he devotes most of his attention to necessity of constitution and considers essentialist principles concerning natural kinds only briefly, neglecting altogether the argument from identity on which I shall focus. In *Reference and Essence*, his concerns are rather closer to mine, but his main target throughout are the views of Hilary Putnam, as expounded in "Is

0022-362X/90/8708/385–398

© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

I

I want to begin by making a distinction between what I shall call honest and underhand essentialist claims; and the scope of my argument can be usefully characterized as vitiating only those which are underhand. The purpose of, and the justification for, the honest/underhand distinction will not be fully apparent until the end of the paper, but I shall attempt a brief characterization here. Very roughly, *honest essentialist claims* are those which derive from recognizably essentialist premises and intuitions, and which therefore leave the anti-essentialist plenty of room for maneuver; for he may simply respond by challenging those premises and intuitions. Into the class of honest essentialist claims must be admitted a very large, important and interesting set of putative necessary truths that are of a kind that I shall label *simply predicative* in form. Examples might be 'necessarily, Aristotle is human'; or 'necessarily, tigers have genetic make-up *G*', where *G* denotes the species-specific genetic code for tigers; or 'necessarily, water is H_2O', where the 'is' here is a simple predicative 'is', i.e., not the 'is' of identity, and 'H_2O', correspondingly, is intended not as a substance name, but as part of a predicate ascribing what might be called H_2O-ness—a chemical property—to water. There is left open to the anti-essentialist in such cases the possibility of a straightforward response. For the establishment of such essentialist truths as these generally proceeds by means of a thought experiment—by appeal to what can and cannot be imagined—and the anti-essentialist is free to question the essentialist's intuitions about the hypothetical cases that he invokes in support of his claims. That is, argument is not at an end when an honest essentialist claim is propounded. The anti-essentialist is still in the game, and has everything to play for.

Semantics Possible?" in H. Kiefer and M. Munitz, eds., *Language, Belief, and Metaphysics* (Albany: SUNY, 1970), pp. 50–63; and "The Meaning of Meaning," in K. Gunderson, ed., *Language, Mind, and Knowledge* (Minneapolis: Minnesota UP, 1975). The following arguments can be seen as a way of directing similar fire toward Kripke's attempts to establish essentialism by means of considerations in the philosophy of language.

It should be noted that Kripke, in the preface to *Naming and Necessity* (New York: Blackwell, 1980), explicitly responds to Salmon's claim that a particular footnote in "Naming and Necessity" [in D. Davidson and G. Harman, eds., *Semantics of Natural Language* (Dordrecht: Reidel, 1972)] suggests an intention to prove essentialism from the theory of reference: "I think I had no ambition in this short footnote rigorously to prove 'essentialism from the theory of reference alone'." It remains unclear whether Kripke supposes that such a proof could be constructed from arguments given elsewhere in his work. At any rate, the argument of "Naming and Necessity" can certainly make it appear as though essentialist conclusions are straightforwardly deduced from compelling considerations in the philosophy of language, and it is this appearance which needs to be countered.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Certain other kinds of essentialist claims, however—notably those expressed by means of the identity relation—appear to present far greater difficulties for the anti-essentialist. These I call *underhand,* and to show why they merit this description is the purpose of this paper.

II

Insofar as the a priori/a posteriori distinction can be allowed to bear weight at all, it is natural to suppose that the discernment of necessity itself must be an a priori matter. This is to salvage from the wreck of the Kantian identification of the necessary with the a priori at least the insight that

> Experience teaches that a thing is so and so, but not that it cannot be otherwise.[2]

It would follow from this that only the truth, and not the necessity, of necessary a posteriori truths could be the subject of empirical discovery. To know *that* a truth is necessary, the Kantian line would have it, requires formal proof or philosophical argument and hence some kind of a priori input.[3]

As far as identity statements are concerned, this suggests a schema of the following form, which I shall call the *identity schema:*

THE IDENTITY SCHEMA

(i) $[(a = b) \rightarrow Nec\ (a = b)]$	Arrived at a priori by philosophical argument
(ii) $(a = b)$	Empirically discovered
(iii) $Nec\ (a = b)$	Overall a posteriori, because of the empical input in (ii), but presenting no epistemological difficulty since the specifically modal knowledge can be attributed to the a priori reasoning at (i)

[2] *Critique of Pure Reason,* Norman Kemp Smith, trans. (New York: Macmillan, 1929), B3–4.

[3] Kripke, in the addenda to *Naming and Necessity,* gives explicit consideration to this question and appears to reject what I have called the natural view: "All the cases of the necessary *a posteriori* advocated in the text have the special character attributed to mathematical statements: Philosophical analysis tells us that they cannot be contingently true, so any empirical knowledge of their truth is automatically empirical knowledge that they are necessary." Kripke puts this forward as a direct counter to the suggestion that Kant might be able to retreat from his wholesale conflation of the necessary with the a priori to a weaker position, holding only that a priori thought is always required to establish that a statement is necessary—that this is something which cannot be empirically discovered. But the suggestion that it is philosophical analysis which permits us to designate certain classes of truths as necessary seems itself to be an expression of the Kantian intuition, and the force of 'automatically' in the above passage is unclear. In particular, it would seem that Kripke would not want to quarrel with the identity schema.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

It is some such picture as this which makes the necessary a posteriori character of certain identity statements seem unproblematic. For it appears to make clear how we may accept the existence of necessary a posteriori truths without flouting the intuitively plausible Kantian suggestion that necessity itself should not be attributable to statements independently of the exercise of some a priori intuition.

As it stands, the identity schema is of only limited scope, for its first premise will not hold for all substituends; in particular, it will not hold when '*a*' or '*b*' is replaced by a definite description. It is usually held, for instance, that though Hesperus is in fact the brightest star in the evening sky, it is not necessarily the brightest star in the evening sky, since there are possible worlds in which other stars are brighter. Strictly, then, in order to derive the conclusion that any particular identity statement is a necessary truth, it is not sufficient that one know the truth of that identity statement, together with schematic principle (i). One must know also that that schematic principle in fact applies to the identity statement in question.

This demand is usually met by means of the constraint that the terms flanking the identity sign be rigid designators. It is easy to see how the constraint renders the schema workable. If '*a*' in fact refers to the same object as '*b*', and if both '*a*' and '*b*' refer consistently to the same object in every possible world as that to which each refers in the actual world, there will be no possible world in which '*a*' and '*b*' refer to different objects. The identity statement formed from placing the terms '*a*' and '*b*' on either side of the identity symbol will thus be a necessary truth. To take account of the rigid designator constraint, we might then amend the identity schema as follows:

(i) For any rigid designators 'x' and 'y': $[(x = y) \rightarrow Nec\ (x = y)]$
(ii) $a = b$
(iii) 'a' and 'b' are rigid designators.
(iv) $Nec\ (a = b)$

I shall return to this schema shortly.

III

I want to turn now, though, to the kinds of essentialist claims with which this paper will be primarily concerned, i.e., identity statements involving natural-kind terms such as 'gold', 'water', etc.[4]

[4] It is often pointed out that such terms—not only natural-kind terms, incidentally, but many common nouns and some adjectives—have a kind of dual usage; that is, they appear to do duty both as names, for example of substances, as in 'Gold is a yellow metal', and as predicates, as in 'This liquid is water' or 'This pan is aluminium'. This is of quite some significance, for it gives rise to the possibility of a number of different ways of formulating underhand essentialist claims: as identity statements in which the identity sign is flanked by substance names; as identity state-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

'Water = H_2O' is a much discussed example. According to Kripke, this provides an example of a true identity, indisputably discovered a posteriori, the terms of which are both rigid designators.[5] It is a simple step, via the amended argument schema above, to the essentialist conclusion that necessarily, water = H_2O. Since the argument concerned will be so frequently referred to in what follows, it will be useful to spell it out now. I shall simply call it the *water argument:*

(i) For any rigid designators, 'x' and 'y': $[(x = y) \rightarrow Nec\ (x = y)]$
(ii) Water = H_2O
(iii) 'Water' and 'H_2O' are rigid designators.
(iv) Therefore, *Nec* (Water = H_2O)

This would seem to be a straightforward derivation of a significant essentialist truth from the philosophy of language, which furnishes us with (i) and (iii), and the empirical result that water is in fact H_2O. Importantly, it is supposed to be easy, on this account, to accommodate the demand that any reasonable account must leave space for the fallibility of science, and so must refrain from insisting on the epistemological certainty of the chemical result that water = H_2O. For it can be admitted by the essentialist that it is just conceivable, albeit fantastically unlikely, that chemistry is in error about this, and that the chemical composition of water is in fact something different. His claim is simply that, *if* water = H_2O, then that is necessarily so. In terms of the water argument, that is, the essentialist can be seen as making the derivation of (iv) depend, very properly, upon the truth of (ii).

In what follows, I want to try to describe a possible world (which is, incidentally, an epistemically possible one, in the sense that, for all we know, it could be the actual world) in which it is not true that water = H_2O. If this thought experiment succeeds, then, since it describes a possible situation in which water is not identical with H_2O, the conclusion of the water argument must be false. Assuming, as I shall, that the argument is valid, we can account for the falsity of the conclusion only by discarding (i), (ii), or (iii). I shall suggest that it is (iii) which has to go.

---

ments in which it is flanked by property terms; and possibly also as some kind of biconditional. I think it is sometimes thought that, while it is clearly a mistake to endorse essentialist claims of the first kind, one can escape difficulties by reformulating the claims as property identities or biconditionals. I hope to show elsewhere that this is not the case. In this paper, though, I restrict my attention to identities involving substance names.

[5] See, e.g., *Naming and Necessity*, p. 140: "Theoretical identities, according to the conception I advocate, are generally identities involving two rigid designators and therefore are examples of the necessary *a posteriori*."

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

IV

Consider the following imaginary case. A certain group of samples of some particular compounds is found not to fit with current chemical theory, in that these samples are found to have significantly higher boiling and melting points than standard samples of those same compounds, though of apparently equal purity, and though this cannot be explained by, e.g., isotopic variations. Investigation results in a most exciting discovery; the anomalous samples are discovered to contain some molecules, the nuclei of whose atoms contain a previously unknown and extremely rare particle, very similar to the proton, and named proton-B. The attraction of proton-B for surrounding electrons is slightly stronger than that of an ordinary proton, and this is what explains the variations in melting and boiling points. It is also discovered, by experiment, that ordinary protons may be readily replaced in the nuclei of all sorts of atoms by these protons-B, and this, not surprisingly, results in certain sorts of changes in properties. We might imagine, for example, that H_2O varies continuously in its properties, depending on how many protons-B are contained in the nuclei of the component atoms in the individual molecules of H_2O. Perhaps in the extreme case, where all the molecules concerned contain only protons-B, H_2O is an opaque, pink solid. Evidently, the discovery of proton-B necessitates a revolution in chemical theory, and a correspondingly radical revision of chemical classification.

What would happen to the claim that necessarily, water = H_2O in such a case as this? Would we continue to insist that all the range of substances, from the pink and opaque to the clear and colorless, which are chemically composed of H_2O are in fact different kinds of water? Surely not; surely we would use different names for the widely varying kinds of substance produced by the substitution of protons-B for ordinary protons.[6] But it is important to realize that nothing

[6] One might take issue with this by adverting to actual cases of isotopic variation, where radical differences in properties have not prevented us from regarding the isotopes in question as different kinds of the same substance. We consider diamond and graphite, for example, to be different kinds of carbon. But this is importantly different from the water case. For water, because of its overwhelming significance for us, is a substance whose name draws primarily on ordinary, everyday usage for its meaning; 'water' is not, except secondarily, a term of chemistry. For this reason, there are limits to its flexibility. Perhaps it has some; to regard the term 'water' as applicable to ice and water vapor might be considered a minor extension of its primary use, made admissible by the readiness with which conversion from one form to another may be effected. But it is rooted too solidly in our everyday lives to permit of total subversion in the cause of chemistry. I submit that it could not come to be used as a name for a pink, opaque solid, whatever the chemical composition of the latter.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

much hangs on this. If we did decide to speak so, that would hardly vindicate the essentialist. What is important is that all plausibility has gone out of the claim that H_2O and water are the very same substance, once the properties to which that particular chemical constitution gives rise are allowed to vary in the way described. In the possible world envisaged, then, water is not identical with H_2O. And so the claim that this identity is a necessary truth must be false.[7]

It may be tempting for the essentialist to reply at this point that nothing I have said so far presents him with any difficulty. For he was never so reckless as to insist simply that necessarily, water = H_2O. His claim was only that 'water = H_2O' is a necessary truth if it is a truth at all, and all that is shown by the thought experiment above is that that identity might be false. This he has never denied.

This is to invoke what I shall call the *modally committed reading* of the identity statement 'Water = H_2O'. For it seems to me that the

---

'Carbon', on the other hand, insofar as it is a name for something of which diamond and graphite are both forms, is principally a term of chemistry; it is indeed only through chemistry that we have been able to associate the two. (I think that insofar as 'carbon' has a common use, it is associated with the amorphous form, as found in soot and coal.) And it is because its roots are here, in chemistry, that it can tolerate wide variations in observable properties, for the sake of an underlying common feature; it is the name, first and foremost, of an element.

[7] One might try to respond to this example by arguing that the possible world envisaged is not a world in which water is not identical with H_2O, because the pink, opaque solid is not only not water, it is not H_2O either. For replacement of ordinary protons by protons-B, it might be thought, brings about such a radical change in properties that we should no longer be justified in retaining the same chemical name for the resulting substance. But it seems to me that this ignores an important aspect of the actual linguistic functioning of chemical terms like 'H_2O'. In effect, the proposal demands that the reference of such terms be responsive to precisely the same considerations as ordinary common names like 'water'. It simply ignores the fact that there is descriptive complexity in a term like 'H_2O', that it is a shorthand for a chemical structure and not just an arbitrary name, so that while that chemical structure (two atoms of hydrogen to one atom of oxygen, arranged in the particular way characteristic of the water molecule) remains intact, the term must continue to apply. It might be replied, of course, that this is a shorthand for a chemical structure developed to deal with a chemistry in which protons-B do not exist, and that, in the chemistry which would evolve, were such particles to be discovered, we would develop different chemical names for the different substances which would result, since it would simply cease to be useful to regard the pink, opaque solid and the colorless, transparent liquid as different kinds of H_2O. But, firstly, this might not be true—compare my discussion of the diamond-graphite case in note 6. In the resulting chemistry, there might be theoretical reasons to retain the common name. And, secondly, the question whether we would call the pink solid 'H_2O' or not, if we had a completely different system of chemical classification, on which 'H_2O' meant something different, does not seem to me to matter much. What matters, for present purposes, is surely whether the pink solid could be called 'H_2O', given the classificatory system from which that term derives its *present* meaning, and this takes no account of proton type.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

content of identity statements of the 'Water = H_2O' variety is ambiguous, and that it is not clear whether, in accepting simply that an identity statement of this kind is *true*, one has already accepted that it is true in all possible worlds (the modally committed reading), or whether one is thereby committed initially only to the view that these substances are identical in the actual world, and that things could have been otherwise (the noncommittal reading). And the thought experiment above, it might be thought, only really threatens the position of someone adopting the noncommittal reading. Anyone adopting the modally committed reading is, as it were, already protected against the possibility of worlds like the proton-B world by the standard caveat which accompanies all his claims—'necessary, if true'.

It should be clear, though, that this is of no help to a defender of the water argument. The problem with the modally committed reading of the identity statement 'Water = H_2O' is that it takes one beyond what can be sanctioned by mere chemistry. Read in the modally committed way, we no longer have reason, on the basis of science alone, confidently to assert the identity at all. For premise (ii) of the water argument, on the modally committed reading, already asserts the impossibility of the proton-B world, and therefore cannot be regarded as the expression of the content of a straightforward, uncontroversial empirical discovery. The identity itself is now a firmly essentialist claim which the proton-B example, if deemed conceivable, simply falsifies.

If premise (ii) of the water argument stands any chance of being true at all, then, it will have to be understood in accordance with the noncommittal reading. On that reading, the thought experiment does not challenge premise (ii); it does show, however, that the conclusion of the water argument is false. Given that (i) seems obviously true, it is premise (iii) that must come under suspicion.

I do not intend to attack the claim, argued very plausibly by Kripke and Putnam, that the term 'water' is a rigid designator. Possibly it could be challenged, but it does not matter, because what I think the thought experiment shows, very simply, is that 'H_2O' is not—at least, not in the sense required for the water argument to work. This requires some elaboration and so it will be useful at this point to introduce a distinction within the class of rigid designators, between those designators whose rigidity is a purely logical property, deriving solely from the way they function in referring, and those whose rigidity is, as it were, *de facto* only. To see how the distinction works, consider for a moment the following example. It is not implausible to suppose that in certain cases one can provide not only necessary but

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

also sufficient conditions of identity for particular individuals. For example, on the assumption that what matters as far as the identity of trees is concerned is position and identity of original seed, one might suggest that a sufficient condition for identity with tree *T*, which grew from acorn *A* at place *P*, is being a tree that grew from acorn *A* at place *P*.[8] Given that this is so, the definite description 'the tree which grew from acorn *A* at place *P*' will be a rigid designator on the usual definition of that term, for by hypothesis it will refer to the same individual in each possible world in which such an individual exists. For any such tree will be tree *T*. This kind of rigidity I call *de facto* only, because the rigidity here results not from any purely logical property of the referring expression involved, but rather from the fact (if it is one) that the description in question simply happens, by virtue of the metaphysics of the matter, to pick out the same individual in each possible world.

Now, my point can be put by saying that 'H_2O', if it is a rigid designator at all, has only this *de facto* kind of rigidity. If it picks out the same substance in each possible world, that will not be because of its logical properties as a referring expression; it will be because chemical composition is, as a matter of metaphysical fact, the basis of the same-substance relation. Essentially, 'H_2O' functions as a descriptive term; it picks out, in each possible world, that substance which is chemically composed of one part oxygen to two parts hydrogen, arranged in the particular way characteristic of the water molecule. It does not function like a name; if it did, it would be impossible to describe worlds like the proton-B world in which there is H_2O that is not water.

Now, is 'H_2O' *de facto* rigid? What the proton-B example appears to show is that it is not. Assuming that we are not just allowed simply to define the same substance relation in terms of chemical composition—that is, assuming that this is something that has been discovered to be true in this world—what seems to have been imagined is a possible world in which 'H_2O' picks out not just one but a large number of different substances, and, if such worlds as this are really possible, then 'H_2O' is not a rigid designator.

Since it can be very tempting to assume that 'H_2O' just must be a rigid designator, I need to expand on how that can fail to be so and to respond to some likely objections. What makes it seem that 'H_2O' must be a rigid designator is, I think, something like the following thought: since 'H_2O' picks out a single substance in the actual world, and since having the chemical composition H_2O is essential to that

[8] Nothing will turn on whether this is a good account of identity for trees or not.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

substance in the sense that on current chemical theory this is what explains its other important properties, the property of what might be called H_2O-ness and being that single substance could never come apart in any possible world. With respect to this thought, I have two remarks:

First, such a line of thought is only defensible at all on a fairly restricted interpretation of what is to count as a possible world. An important feature of the term 'H_2O' is that it picks out, in every possible world, the substance that has the chemical constitution H_2O in that world, so that, unless we have a watertight guarantee that the H_2O in every possible world is bound to be the same stuff, there will be possible worlds in which H_2O is not the familiar clear, colorless substance we call water. If I can be allowed to invoke the common —though admittedly somewhat dubious—distinction between logical and physical possibility, the point could be put by saying that, even if in fact it is inevitable, according to some relevant laws of nature, that H_2O-ness should give rise to certain physical properties, there are certainly logically possible worlds in which H_2O-ness gives rise to different macroproperties in a substance from the ones to which it actually does give rise—the world of the thought experiment being just one example. And hence it would seem that H_2O does not pick out the same substance in every logically possible world. Doubtless, though, some essentialists would not be particularly put out by this claim; they would simply dispute the intelligibility of the notion of logical possibility invoked, and remain happy with the more limited conclusion that 'H_2O' at any rate picks out the same substance in every physically (or perhaps metaphysically) possible world. Kripke, for instance, talks not of logical but of metaphysical possibility, and mentions at one point that physical necessity might turn out to be "necessity in the highest degree."[9] So it is not clear that the essentialist would be perturbed by my observation. That it is in some sense logically possible that a substance having the chemical composition H_2O should be a pink, opaque solid may not concern him at all. And if 'H_2O' were to be a rigid designator at least with respect to all physically possible worlds, then it would allow one to deduce, by means of an argument structurally identical to the water argument, that it is at least physically necessary that water = H_2O, that is, of course, if it is true at all.

Any hope the essentialist might have had of getting away with this much is vain, however. The problem lies in the fact that any attempt to establish the truth of the relevant version of premise (iii), i.e., to

[9] See *Naming and Necessity*, p. 99.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

establish that 'H_2O' is *de facto* rigid across physically possible worlds, simply requires that the conclusion of the argument, i.e., that necessarily, water = H_2O, be already taken for granted, and so cannot be used in the proof of that conclusion. To make the point clear, consider once more the case of tree *T*, grown from acorn *A* at position *P*. I think it is evident that one would not want to argue to the conclusion that 'necessarily, *T* = the tree grown from acorn *A* at position *P*' by means of the following argument:

(i) For rigid designators '*x*' and '*y*': $[(x = y) \rightarrow Nec\ (x = y)]$
(ii) *T* = the tree grown from acorn *A* at position *P*
(iii) '*T*' and 'the tree grown from acorn *A* at position *P*' are rigid designators.

Therefore,

(iv) *Nec* (*T* = the tree grown from acorn *A* at position *P*)

The reason is that premise (iii), which asserts that 'the tree grown from acorn *A* at position *P*' is a rigid designator (i.e., *de facto* rigid) just does too much work in this particular argument; once it is established that 'the tree grown from acorn *A* at position *P*' is an expression that refers to the same object in every possible world, one already has one's conclusion. That the result has something to do with the necessity of identity and rigid designation is patently not true; all the weight in this argument is being borne by the philosophical intuitions that support premise (iii). To claim that the essentialist conclusion expressed in (iv) can be derived solely from certain plausible theses in the philosophy of language would clearly be absurd.

Now, my claim is that the water argument has exactly the same structure as the argument (i)–(iv) above. Because the rigidity of a term like 'H_2O' can be *de facto* only, the establishment of premise (iii) demands more than reflection upon the behavior of referring expressions; specifically, it demands that one be prepared to assert that 'H_2O' picks out the same substance in every physically possible world in which there is any H_2O. It demands, in other words, that one be prepared to accept the conclusion for which one is supposedly arguing before one has reached it by means of the water argument.

Note that I do not mean to suggest that any such acceptance would be independently indefensible. It is not unreasonable to assume the approximate correctness of the chemical theory upon which current classification of substances is based, and with it a concomitant claim about the inexorability with which chemical composition gives rise to macroproperties. I am not suggesting that it is likely that protons-B exist. What it is important to see is that no argument remains for the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

essentialist's conclusion of the kind that the structure of the water argument suggested might be possible, i.e., an argument importing no substantial metaphysical presuppositions into its premises. For, clearly, the kinds of arguments one would need to give in support of the *de facto* rigidity of 'H_2O' need to rely on the very kinds of metaphysical considerations one might appeal to in attempting to establish that necessarily, Water = H_2O by more direct means.

It is important, incidentally, to note that the possible world described by the thought experiment is not a possible world in which water has been shown not to be H_2O, where the 'is' here is a *simply predicative* 'is'. Water, in the situation described, still has the chemical composition H_2O; it is just that a full chemical explanation of its macroproperties would need to advert to more than this fact, in particular, to its having no proton-B in the nuclei of the atoms of its molecules. Since, however, such a full explanation would still need to advert to water's being H_2O to explain at least a substantial part of those properties, it seems to remain plausible that water might essentially have the chemical constitution H_2O. The point is, that the claim that 'necessarily, Water is H_2O', where the 'is' is a predicative 'is', is an honest essentialist claim, and a good candidate, it would seem, for a true one. The thought experiment I invoked has no bearing on the question whether this honest claim is correct.

The point about the underhand identity claim is that it combines with the predicative claim at the very least the additional claim that necessarily, H_2O is water; for the identity relation is symmetrical. Incidentally, I think the strangeness of the sound of this latter claim is an indication that most of the time we tacitly read 'is' in place of the identity sign, and tacitly assume that the 'is' here is a simply predicative 'is'—perhaps one reason why it is usually thought that Twin-earth considerations and the like are sufficient to establish the identity claim. But that is not the case; the most that could possibly be claimed for such thought experiments is that they show that necessarily, water is (predicatively) H_2O. Importantly, they do not show that necessarily, H_2O is water, and that leaves room for the possibility of worlds like the proton-B world just described.

V

Although my prime concern has been to establish that arguments like the water argument do not proceed, as may at first appear, from metaphysically innocuous premises to substantial essentialist conclusions, it may be interesting to consider, briefly, what may be said independently in support of the conclusion that it is necessary, in some limited sense (perhaps physically necessary), that

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

water = H_2O.[10] The question under consideration is whether there is a *physically* possible world in which the stuff that has the chemical composition H_2O in that world is not the same substance as the substance we call water in our world. Is there such a world? I think the answer to this is simply that we do not really know. To insist that there is no such world is to insist that current chemical theory is correct in what it postulates to be the ultimate explanatory basis of the macroproperties of substances; it is, in other words, to rule out by fiat revolutions in chemical theory of the kind envisaged by the proton-B example. But this is surely not something we can be confident in asserting. To make the point vivid, recall that the world of the thought experiment is an epistemically possible world, in the sense that, for all we know, it is a world that the actual world could turn out to be! There might *be* protons-B, or other comparable theoretical developments lurking in the wings, to upset the whole basis of chemical classification and theory. We just cannot rule that out a priori. In the terms of the previous discussion, this could be put by saying that we just do not know whether 'H_2O' is *de facto* rigid across physically possible worlds.

What is the effect of this ignorance on the claim that it is physically necessary, if true, that water = H_2O? We now find that anybody wishing to support such a claim will have to qualify further his claim to shield it from the possibility of overthrow in the face of empirical advance; he will have to say not only that his claim is conditional on the truth of the noncommittal reading of the identity itself, but also on the correctness of the classificatory principles presupposed by current chemical theory. It is conditional, that is, upon the assumption that chemical composition at the molecular level is in fact the ultimate explanatory basis for the individuation of substances. And that, in turn, is to presuppose that it is physically impossible that H_2O-ness should give rise to any macrostructural properties other than the ones to which it does in fact give rise; or, which is the same thing, to presuppose that 'H_2O' is a *de facto* rigid designator across physically possible worlds.

But when thus made conditional upon the truth of a premise concerning the correctness of the chemical theory on which current classification of substances is based, the essentialist's claim lapses into triviality. What he now appears to be saying is this: provided it is

[10] I am taking it to have been already proven by the proton-B example that, once one interprets the notion of possibility in its usual wide sense, it is simply false that it is impossible for water not to be the same substance as H_2O.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

true that water is actually H_2O, and provided it is true that anything which is H_2O is, as a matter of physical necessity, a substance identical with water in its properties, then it is physically necessary that water is identical with H_2O. This is a claim of the form: 'If p, and if necessarily p, then necessarily p'. And surely it is not only the essentialist who will be happy with that.

I have argued, then, that the claim that identity statements between substance names are necessary truths, in any strong sense, is a false claim, and that arguments on the model of the identity schema in support of such claims rest on the false premise that both terms involved are true rigid designators. I want to conclude with a few remarks about the bearing of what I have said on the claim that the kind of statement I have discussed provides examples of the necessary a posteriori. The upshot of my arguments is, I think, this: that while such identity statements may indeed be necessarily true in some weak sense of 'necessarily', we cannot know them so to be; and the reason for this has nothing to do with our not being entirely certain of their truth. And so a 'necessary, if true' clause will not do to defend the essentialist against the charge that he has assumed that the science relevant to the area of his claim has reached its zenith. What he needs, in effect, is a 'necessary, if necessary' qualifying clause, which destroys his claim to have come up with a result of any significance.

Indeed, it is entirely possible that we shall never know whether identity statements of the 'water = H_2O' variety are necessarily true, however much science progresses.[11] A clear insight into the inexorability with which certain combinations of properties give rise to others may be something it will never be our privilege to possess. For however far science goes in the direction of explaining phenomena in terms of more fundamental states and properties, we shall always be forced back, at some stage, to a position where we can say nothing more than that this is how it is. The shift to 'this is how it must be' is tempting, but it is hard to see what could ever justify it.

HELEN STEWARD

St. Hugh's College/Oxford

[11] Again, it is only a weak sense of 'necessarily' that is under consideration.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

WHAT IS REASONING? WHAT IS AN ARGUMENT?*

THESE are changing times for logic, as the subject is studied and taught in philosophy departments. In recent years, there has been the advent of informal logic as a growing phenomenon in teaching logic, the demand for critical thinking in education, and increased questioning of the role and usefulness of formal logic. There has also been the remarkable growth of the newly founded interdisciplinary field of argumentation.

In light of these developments, it has become questionable whether there are any longer clear and appropriate definitions of the key terms 'reasoning' and 'argument' with which we can work. Charles Hamblin's[1] pioneering chapter seven, "The Concept of Argument," articulated the nature and importance of the problem with foresight. But as he clearly stated, Hamblin chose to circumvent attempting to address it directly, confining his treatment to the question of what a good (successful) or a bad (fallacious) argument is:

> There is little to be gained by making a frontal assault on the question of what an argument is. Instead, let us approach it indirectly by discussing how arguments are appraised and evaluated (*ibid.*, p. 231).

Although there have been significant developments in the area of informal logic since 1970,[2] the failure to arrive at any clear agreement on, or analysis of, the meaning of the concepts of argument and reasoning is a fundamental difficulty.

A fallacy is often said to be an erroneous argument or an instance of faulty reasoning.[3] But which is it? Or are both characterizations accurate? It seems hard to say, no doubt partly, at least, because of the generality, vagueness and heterogeneity of usages associated with the key terms 'argument' and 'reasoning'.

Rob Grootendorst[4] advocated, "let's stop restricting ourselves to looking upon fallacies as errors of reasoning," because this approach

* Research for this paper was supported by three awards: (1) a Killam Research Fellowship awarded through the Canada Council; (2) a Fellowship from the Netherlands Institute for Advanced Study in the Humanities and Social Sciences; and (3) a Research Grant from the Social Sciences and Humanities Research Council of Canada. I would like to thank Frans van Eemeren, Rob Grootendorst, Sally Jackson, Scott Jacobs, Agnes Haft-van Rees, Francisca Snoeck Henkemans, Eveline Feteris, Erik Krabbe, Agnes Verbiest, Charles Willard, and John Woods.

[1] *Fallacies* (London: Methuen, 1970).

[2] Ralph H. Johnson and J. Anthony Blair, "Informal Logic: The Past Five Years 1978–1983," *American Philosophical Quarterly*, XXII (1985): 181–196.

[3] See Hamblin, and my *Informal Fallacies* (Amsterdam: Benjamins, 1987).

[4] "What a Pragma-Dialectical Approach to Fallacies Can and Cannot Do," paper read at the Third International Conference on Informal Logic, June 1989.

0022-362X/90/8708/399–419
© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

trivializes them to "uninteresting mistakes in logical reasoning." Instead, Grootendorst proposes viewing fallacies as bad arguments in the sense of being Gricean failures of co-operation which violate rules of a critical discussion. This highly sensible proposal leaves room for the idea that critical discussion is or can be, however, a reasonable form of discussion because it (in some sense) contains reasoning. Even if we concede, therefore, that fallacies are failures of argument, it could still be true that they are (in some yet to be clarified sense) also failures of reasoning. But, of course, none of this makes any sense unless we can say, in general, what the difference is between argument and reasoning.

Internalists, like Gilbert Harman,[5] have portrayed reasoning as a mental, psychological, or internal process, "a procedure for changing one's view" (*ibid.*, p. 107). Externalists, like Jim Mackenzie,[6] have portrayed reasoning as a process of linguistic interaction that appears to be more sociological than psychological.

Argument has stereotypically been portrayed by logic textbooks as an externally manifested set of propositions "designated" as premises and conclusion. But often, at least in initially describing what an argument is, the texts do some hand waving to the effect that the conclusion is a "claim" based on "reasons" given in the premises (but little is made of this subsequently, in many of the texts). The other point of view on argument, a sharply contrasting one, is the pragma-dialectical approach of Frans H. van Eemeren and Grootendorst[7] which emphasizes public commitments over personal beliefs, and sees an argument not as an internal, mentalistic process but as a rule-governed kind of discussion entered into by two or more parties in order to resolve a conflict of opinions. According to this much broader approach, an argument is more than just a set of propositions. It comprises many kinds of speech acts, evaluated in a goal-directed, normative model of dialogue.

Given these various points of view, which now seem to be on the brink of changing new developments in logic and the study of reasoning and argument generally, it is problematic to see how reasoning is related to argument. Are reasoning and argument essentially the same thing? Or is one a proper subpart of the other? Or can you have reasoning that is not in argument? Or could you have argument without reasoning? It seems hard to know where to begin replying to these questions.

[5] "Logic and Reasoning," *Synthese*, LX (1984): 107–127.
[6] "Reasoning and Logic," *Synthese*, LXXIX (1989): 99–117.
[7] *Speech Acts in Argumentative Discussions* (Dordrecht: Foris, 1984).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Certainly these questions have begun to bother those of us working in the area of informal logic and, from time to time, stabs at saying something useful have been made. Trudy Govier[8] has made a highly plausible characterization of the basic difference between argument and reasoning as follows.

> An argument is a publicly expressed tool of persuasion. Typically it takes thinking to construct an argument. Reasoning is distinguished from arguing along these lines: reasoning is what you may do before you argue, and your argument expresses some of your (best) reasoning. But much reasoning is done before and outside the context of argument (*ibid.*, p. 117).

The account given below will be in basic agreement with Govier in seeing reasoning as something that takes place in argument, and will portray argument as a social, interactive, goal-directed tool of persuasion. Many other things, as well as persuasion, will be central, however.

In the account of reasoning and argument now to be presented, reasoning will be defined as a kind of abstract structure, which can nevertheless be dynamic and interactive in some cases, as well as static and solitary in other cases. In this account, reasoning is characteristically used in argument, but it can be used in other pragmatic contexts as well.

I. WHAT IS REASONING?

In his *Dictionary of Philosophy,* Peter A. Angeles[9] defines *reasoning* in three ways.

> reasoning. 1. The process of inferring conclusions from statements. 2. The application of logic and/or abstract thought patterns in the solution of problems or the act of planning. 3. The ability to know some things without recourse directly to sense perceptions or immediate experience (*ibid.*, p. 240).

The second definition is not too helpful until we know how "logic and/or abstract thought patterns" are being defined or understood. For example, is formal logic or informal logic meant? The third definition is a negative one that excludes resource to sense perceptions. The restriction seems appropriate and instructive, but it does not offer a positive account of what reasoning is.

The first definition is positive, and does seem to capture the basic idea behind reasoning. The main problem with it is the use of the

[8] "Critical Thinking as Argument Analysis?" *Argumentation,* III (1989): 115–126.

[9] New York: Barnes & Noble, 1981.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

term 'inferring'. Can this term be defined without presupposing our understanding of it as a kind of reasoning? Perhaps it can, if we define an *inference* as the use of a rule or warrant to link some propositions (statements) with others. The *conclusion* is the proposition toward which the inference moves. The *premises* are the beginning propositions from which the inference starts. Thus, an inference links the premises to the conclusion, and it always has a direction—proceeding from the premises to the conclusion.

Reasoning is primarily from propositions to propositions, but the conclusion of reasoning can also be an action. In such a case, however, the action is within an imperative that directs an agent to do something (or to abstain from doing something). This kind of imperative could also be described as a practical ought sentence of the form: "Agent, you ought to do such-and-such!" While, strictly speaking, not a proposition, this kind of sentence can behave somewhat like a proposition in practical reasoning, because it can follow from premises that describe an agent's goals and knowledge of his particular circumstances.[10] It is a practical ought sentence that expresses an imperative.

Another valuable feature of Angeles's definition of 'reasoning' is that it rightly states that reasoning needs to be understood within a context of use.

> Reasoning can be used for a variety of purposes: to deceive, to argue, to debate, to doubt, to persuade, to express, to explain, to apologize, to rationalize, etc. It seems that any form of conscious activity can be affected and structured by the reasoning process (*op. cit.*, p. 240).

Note that 'argument' and 'reasoning' are conceived here as different (nonequivalent) terms. Reasoning is used in argument. Reasoning can be used in different speech acts, or contexts of discourse.

Logical pragmatics could be conceived of as the study of the uses of reasoning in a context of discussion. But what, then, is logic? Is it the study of the abstract forms of reasoning exclusively, or can it comprise the uses of reasoning, as the emerging practices of so-called informal logic seem to suggest? One thing we can say is that in a critical discussion,[11] logical reasoning can be used where one party, in dialogue with another party, tries to convince this other party that his (the first party's) point of view is right. In this use of reasoning, the premises are concessions made by one party to the discussion, typically at the request of the other party. Conclusions are proposi-

[10] Cf. B. J. Diggs, "A Technical Ought," *Mind*, LXIX (1960): 301–317.
[11] Van Eemeren and Grootendorst, *op. cit.*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tions drawn from these premises by the warrants appropriate for the discussion, or agreed to by the participants.

Along these lines, then, it seems appropriate to offer an abstract definition of reasoning which, in order to be fully understood, needs to be placed in a pragmatic context of use. As a first pass, let us consider the following proposed definition.

Reasoning is the making or granting of assumptions called *premises* (starting points) and the process of moving toward conclusions (end points) from these assumptions by means of warrants. A *warrant* is a rule or frame that allows the move from one point to the next point in the sequence of reasoning. The term 'warrant',[12] used instead of the more familiar (but narrower) term 'rule', is appropriate because of the existence of frame-based, and other kinds of non-rule-based reasoning.

This definition appears to be comparable to the one given by Dagobert D. Runes[13] in his *Dictionary of Philosophy*.

> Reasoning is the process of inference; it is the process of passing from certain propositions already known or assumed to be true, to another truth distinct from them but following from them; it is a discourse or argument which infers one proposition from another, or from a group of others having some common elements between them (*ibid.*, p. 281).

The main difference between our definition (which is closer to Angeles's definition in this respect) and Runes's definition is that Runes defines reasoning as a discourse or argument, whereas we define reasoning as occurring within discourse or argument.

A valuable aspect of Runes's approach to defining 'reasoning' is the careful division of the definition into two parts or meanings. The part quoted above refers to reasoning *in logic*. Another quite different part defines reasoning *in psychology* as "the act or process of exercising the mind," referring to actual thought processes (*ibid.*, p. 280). By contrast, Runes sees the definition of reasoning in logic as being essentially normative in nature.

> It must be added that the logical study of reasoning is normative: logic does not analyze it simply in its natural development, but with a view to guide it towards coherence, validity or truth (*ibid.*, p. 281).

Reasoning can be studied from various points of view. Psychology, sociology, and other empirical sciences study the behavioral dynamics of how reasoning is actually carried out by individuals or

[12] Stephen Toulmin, *The Uses of Argument* (New York: Cambridge, 1969, 2nd ed.).
[13] Totowa, NJ: Rowman & Allanheld, 1984.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

groups. To study the logic of reasoning implies a normative point of view whereby the reasoning is judged as weak or strong, good or bad, valid, fallacious, etc. With the rise of mathematical logic as a discipline, a sharp wedge was driven between the logical and the psychological points of view on studying reasoning. While there are overlapping concerns and interests, still it is important to see clearly that the logical point of view on reasoning is distinctively different from the psychological point of view.

From the logical, as opposed to the psychological point of view, *reasoning* can be defined generally as a sequence of steps from some points (premises) to other points (conclusions). Reasoning is sequential, and best modeled abstractly by a graph or argument diagram where arcs (steps) link points (vertices in the graph).[14] The points normally represent propositions, but they can represent the contents of other speech acts, like questions and imperative utterances, in some cases.

Reasoning can be linear, or it can have a tree structure, or it can even be circular in some cases (*ibid.*).

Reasoning normally has a direction. Most often, as already mentioned, it moves "forward" from the premise toward the conclusion. Directionality of reasoning depends, however, on how it is being used in a context of argument. There can be "backward" reasoning in some instances—for example in a kind of case where a conclusion is known and the reasons (premises) supporting the conclusion are sought out. Backtracking of this sort is not always reasoning, but it can be a kind of reasoning in some cases.

In the context of argumentation characteristic of "proving your point" in a critical discussion, reasoning goes forward from your opponent's premises to your own conclusion. Forward reasoning is generally characteristic of cases where reasoning has a *probative function,* that is, where the premises are being used to support the conclusion, or to prove it. The directed graph (digraph) is the model of directional reasoning.

Reasoning is normally directed, in a context of argument, toward some goal. But it need not always be. There can be aimless reasoning. When reasoning is used to fulfill a probative function, there is normally one particular proposition designated as the ultimate conclusion to be proved in the sequence of reasoning. This goal gives the sequence of reasoning its purpose. The reasoning should be judged as good, appropriate, erroneous, fallacious, etc., insofar as it supports the realization of this goal, or prevents it from being fulfilled.

[14] Walton and Lynn M. Batten, "Games, Graphs and Circular Arguments," *Logique et Analyse,* LVI (1984): 133–164.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

II. KINDS OF REASONING

There are four especially important categories that define different kinds of reasoning.

1. Monolectical and dialectical reasoning: *dialectical (interactive) reasoning* occurs where there are two participants reasoning together, and the reasoning of each participant contains steps derived from the reasoning of the other. *Monolectical reasoning* is nondialectical, i.e., a single reasoner can function alone, requiring no input from another reasoner.
2. Alethic and epistemic reasoning: *alethic reasoning* is based only on truth values of propositions, whereas *epistemic (knowledge-based) reasoning* takes place in relation to a knowledge base. In knowledge-based reasoning, a reasoner can draw on a given set of premises that are known to be true.
3. Static and dynamic reasoning: in *static reasoning*, the set of given premises is fixed or designated, and cannot change. In *dynamic reasoning*, this set can change at each succeeding step in reasoning. Dynamic reasoning is relative to (open) changeable circumstances.
4. Practical and theoretical reasoning: *practical reasoning* is a kind of goal-directed reasoning that seeks out a prudential line of conduct for an agent in a particular situation.[15] *Theoretical (discursive) reasoning* seeks evidence that counts for or against the truth of a proposition. The primary characteristics of practical and theoretical reasoning are summarized below.

Theoretical reasoning is oriented to finding reasons for accepting a proposition as true or false. The problem is to decide whether a proposition is justified on the basis of what we know: *cognitive orientation.*	*Practical reasoning* is oriented to choosing a course of action on the basis of goals and knowledge of one's situation. The problem is to decide whether an action is practically reasonable or prudent: *practical orientation.*
Premises are assumptions that support, or fail to support, a conclusion to a particular extent. There are three standards of support: 1. deductive standards; 2. inductive standards; 3. plausible (opinion-based) standards.	Practical reasoning is characteristically based on uncertainty or incomplete knowledge of a particular (changing) situation. Premises describe goals and knowledge. The conclusion is an imperative.

[15] See G. E. M. Anscombe, *Intention* (New York: Blackwell, 1957); and G. H. von Wright, "On So Called Practical Inference," *Acta Sociologica*, XV (1972): 39–53.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Theoretical and practical reasoning are vehicles for solving different kinds of problems. Hence the methods used are different, even though overlapping. Practical reasoning characteristically arises in a situation that is a *practical conflict,* where, no matter what an agent does, he violates one or the other of his commitments. An example is the situation of Antigone who was committed by divine law to bury her slain brother Polyneices, but also committed as a citizen of Thebes to obey the king's order that no one was to bury Polyneices.

The comparable concept of theoretical reasoning is *logical inconsistency,* where there is a set of propositions that are logically inconsistent, meaning that it is not logically possible (i.e., semantically possible) for all to be true.

A final important distinction is that in all reasoning, there can be an open-world assumption or a closed-world assumption. The *closed-world assumption* means that the reasoning is treated as static, and the premises are regarded as exhaustive of all the evidence relevant to the conclusion. This assumption says: "This is all we need to know, and no further evidence will count as relevant." In any reasoning where the closed-world assumption does not obtain, then the *open-world assumption* obtains, meaning that there is the possibility of new information affecting the reasoning.

The open-world/closed-world distinction is crucial to evaluating lack-of-knowledge inferences, or arguments traditionally called *argumentum ad ignorantiam,* of the following type. Let A be a proposition.

(IG1) If A is not known to be true, then A is false.
(IG2) If A is not known to be false, then A is true.

In cases of reasoning where the closed-world assumption obtains, (IG1) and (IG2) are very strong (indeed, deductively valid) forms of reasoning. In cases where the open-world assumption obtains, (IG1) and (IG2) can be stronger or weaker, depending on the particular case. It follows that the *argumentum ad ignorantiam* is (a) not always fallacious, and (b) depends on features of the reasoning in a particular case as an argument that is stronger or weaker.

III. PRACTICAL REASONING

Practical reasoning is a kind of goal-directed reasoning based on two premises. One premise states that an agent has adopted a goal or intention. The other premise states that this agent knows that, relative to his given situation, there is an action that is a means to carry out the goal of the first premise. The conclusion of a practical inference states that the agent ought (practically) to carry out the action cited in the second premise.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Practical reasoning is a chaining together of the two basic schemes of practical inference represented below, where *a* is an agent, *A* is an action, and *G* is a goal.[16]

G is a goal for *a*.
Doing *A* is necessary for *a* to carry out *G*.
Therefore, *a* ought to do *A*.

G is a goal for *a*.
Doing *A* is sufficient for *a* to carry out *G*.
Therefore, *a* ought to do *A*.

The first type of practical inference is called the *necessary-condition scheme*, and the second type is called the *sufficient-condition scheme*. Practical reasoning is used in different contexts of argumentation, and may be judged as successful or correct in relation to the goals of these contexts.

One such context is planning. Planning arises when an agent is considering possible alternative future courses of action as ways of confronting a conflict of commitments or uncertainty on how to proceed in a situation where some action is called for.

The concept of a plan has achieved recognition in AI as a fundamental in script-based reasoning (reasoning based on common-sense knowledge of familiar types of situations). Robert Wilensky[17] has presented a plan-based theory of understanding as a tool for studying problem-solving and natural-language reasoning for AI. The special aspects of planning highlighted in Wilensky's study are multiple goal planning, resolution of goal conflicts, and the reasoned process of goal abandonment. The kind of goal-directed knowledge structure central to Wilensky's concept of planning is clearly a framework of systematic, intelligent deliberation which provides one important context for practical reasoning.

Another context of practical reasoning (often associated with or included in planning) is that of advice-giving dialogue where an expert in a skill or domain of knowledge is consulted in order to solve a problem. Generally, practical reasoning is best judged as argumentation within a framework or context of interactive argument as a shifting of a burden of proof from one participant to another.

For each argumentation scheme, of either of the two kinds of schemes above, there is a matching set of critical questions of the following kinds.

[16] See my *Practical Reasoning: Goal-Driven, Knowledge-Based, Action-Guiding Argumentation* (Savage, MD: Rowman & Littlefield, 1989), p. 48.

[17] *Planning and Understanding: A Computational Approach to Human Reasoning* (Reading, MA: Addison Wesley, 1983).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

 Digitized by Arya Samaj Foundation Chennai and eGangotri

1. Are there alternative means of realizing G, other than A?
2. Is it possible for a to do A?
3. Does a have goals other than G, which have the potential to conflict with a's realizing G?
4. Are there negative side effects of a's bringing about A that ought to be taken into account?

If the two premises of a scheme are satisfied by a proponent, then a burden of proof is thrown onto a respondent who questions the conclusion based on these premises. The burden implies that the respondent should pose one or more of the appropriate critical questions above. If he does so, the burden is then shifted back onto the proponent to reply adequately to the question.

Practical reasoning is the vehicle whereby stated or formulated goals are linked as commitments to specific actions relative to the particular situation of an agent (as the agent sees the situation).[18] Practical reasoning is the chaining together of practical inferences, bringing together the commitments of an agent in relation to the agent's situation.

Thus, practical reasoning, in contrast to theoretical reasoning, is a feedback relationship between an intelligent agent (knowledge-base) interacting with its changing external circumstances in order to guide itself toward a goal.

IV. WHAT IS AN ARGUMENT?

In his *Dictionary of Philosophy,* Angeles gives a two-part definition of *argument.*

> argument (L., *arguere,* "to make clear"). **1.** The reasons (proof, evidence) offered in support or denial of something. **2.** In logic, a series of statements called *premises* logically related to a further statement called the *conclusion* (*op. cit.,* p. 18).

Part 2 of this definition falls back onto part 1, for presumably the *premises* are the "reasons (proof, evidence) offered in support or denial of" the *conclusion.* Offering a helpful definition of *argument* then requires an elucidation of what are reasons (proof, evidence) offered in support (or denial) of something.

Definition 1 above is the standard sort of definition of 'argument' given in logic textbooks. But there are variants of it. Irving Copi[19] requires only claims "which are regarded as providing support," rather than genuine reasons.

> An argument, in the logician's sense, is any group of propositions of which one is claimed to follow from the others, which are regarded as

[18] Cf. *Practical Reasoning.*
[19] *Introduction to Logic* (New York: Macmillan, 1986, 7th ed.), p. 6.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

providing support or grounds for the truth of that one. Of course, the word "argument" is often used in other senses, but in logic it has the sense just explained (*ibid.*, p. 6).

Claimed by whom? And regarded by whom? By the proponent of the argument, one would suppose.

With respect to Angeles's definition, do the reasons (proof, evidence) have to be good reasons? Or can there be bad or erroneous reasons offered in support of something? Can there be erroneous proof or wrong evidence offered in support of something, yet where, even so, that would be enough proof or evidence to call what was offered an argument? It would seem that a proponent of the Angeles definition should allow for these possibilities. Certainly, it should not follow from a definition of 'argument' that all arguments are good arguments, and that there are no bad arguments.

This brings us back to Copi's definition. What is meant by 'claimed' here, and 'regarded as providing support'? Evidently these phrases refer to a kind of stance or attitude taken up or conveyed by the proponent of the argument. To claim that a proposition is true and can be supported is to assert that proposition and commit oneself to its truth, implying a commitment to defending its truth, as alleged, against attacks or undermining of it by any potential opponent. In this sense, the term 'claim' tacitly presupposes an interactive (dialectical) framework of a proponent upholding a point of view and an opponent questioning that point of view. A claim is an upholding of some particular proposition that is potentially open to questioning.

Copi's definition, however, only goes part way toward the dialectical conception of an argument. In this regard, it is typical of the logician's use of the term 'argument' in logic texts and manuals since Aristotle, where there is the attempt to suppress the idea of an interactive context of discussion. The perceived need is to see the concept of argument as a purely objective notion that can be captured by the formal logic of propositions and truth values. In this standard approach, the dialectical meanings of the term 'claim' are suppressed, and never again mentioned.

Among those not corrupted by logic courses, however, the term 'argument' has a broader meaning. According to *Webster's New Twentieth Century Dictionary of the English Language*,[20] 'argument' is given three basic senses (disregarding the technical and obsolete senses also listed).

ärˈgū-ment, *n.* [Fr. *argument;* L. *argumentum,* evidence, proof, from *arguere,* to make clear, prove.]

[20] New York: Rockville, 1965.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

1. A reason offered for or against a proposition, opinion, or measure; a reason offered in proof, to induce belief, or convince the mind; as, the only *argument* used was force.

2. A debate or discussion; a series of reasoning; as, an *argument* was had before the court, in which *argument* all the reasons were urged.

3. The subject of a discourse or writing; in a more restricted meaning, an outline of the plot or a summary of any literary production; as, the *argument* of a play.

The abstract or *argument* of the piece is shortly as follows.

The first meaning is similar to the definitions of Angeles and Copi, except that it is broader, including the purposes of inducing belief or convincing the mind. These purposes appear to be psychological in nature, but could perhaps alternatively be construed as dialectical elements, i.e., a proponent having the purpose of convincing an opponent in an argumentative discussion.

But the dialectical aspect is explicitly expressed in the second meaning. According to this broader meaning, the argument is not identical to the reasons offered to support a conclusion. Rather, the reasons are advanced *in* the argument.

The third meaning also expresses a broad notion of argument. In this sense, an argument is not just a localized step of inferences, consisting of a small number of premises and a conclusion, connected by a single warrant. Instead, it is a long thread or fabric that runs through and holds together an extended discourse or argumentative text. In this sense, the argument could be expressed in a summary, but the full argument could have originally been a linked sequence of subarguments, ranging over an entire book for example.

The Webster's entry suggests a conception of argument that is much broader and richer than the truncated version of it partially expressed in the traditional logicians' definition. This broad notion is more like what van Eemeren and Grootendorst call *argumentation,* a goal-directed form of interactional (communicative) activities wherein two parties attempt to resolve a conflict of opinions (*op. cit.*, pp. 1–9). In the *simple critical discussion,* one party (the proponent) has the role of defending an expressed opinion (point of view), and the other party has the role of critically questioning that opinion.

Such a pragmatic perspective suggests a new way of defining 'argument' to make it coextensive with argumentation. The only difference between the two is one of connotation. 'Argumentation' refers to the global process of defending and criticizing a thesis (point of view) which spans the whole context of discussion. The term 'argument' can also have this meaning, but is often used for practical purposes to refer to a local segment of a chain of argument, com-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

prising specifically designated premises and conclusions. According to this usage, the term 'argument' can be used in a restricted way somewhat reminiscent of the logicians' truncated definition. Accordingly, a new definition of 'argument' is now proposed.

Argument is a social and verbal means of trying to resolve, or at least to contend with, a conflict or difference that has arisen or exists between two (or more) parties. An argument necessarily involves a claim that is advanced by at least one of the parties. In an asymmetrical case, one party puts forward a claim, and the other party questions it. In a symmetrical case, each party has a claim that clashes with the other party's claim. The claim is very often an opinion, or claim that a view is right, but it need not be. In a negotiation argument, the claim could be to goods or to financial assets.

The conflict or difference (*stasis*) that is the origin of the argument could be of different kinds—it could be a conflict of opinions, an unsolved problem, an unproven hypothesis, or even a situation where both parties are blocked from further actions they are trying to carry out. The different kinds of argument are different ways of trying to resolve these conflicts.

One of the most familiar and important kinds of argument is the persuading- (or convincing-) oriented kind called the critical discussion by van Eemeren and Grootendorst. But there are other kinds as well.

V. ARGUMENT AND DIALOGUE

Reasoning normally occurs in a framework of use (pragmatic framework). Often, the framework of use is argument. Reasoning does not necessarily or always occur in argument, however. A participant can reason in a game of chess, for example, where the reasoning need not necessarily be in an argument. Or to take another example, reasoning can occur in offering or understanding an explanation, where the reasoning is not in an argument.

When reasoning occurs in a context of argument, we say, derivatively, that there are different kinds of reasoning. This way of speaking is perfectly intelligible and acceptable, but it should be realized that such differences are not intrinsic to the reasoning. Rather, they are different kinds of reasoning only in the derived sense that the reasoning is occurring in a different context of argument, meaning that it is being used differently.

Reasoning can be aimless, but argument is essentially goal-directed. And so when reasoning occurs in a context of argument, that reasoning is purposive. Just as reasoning occurs in argument, argument occurs in a larger context of activity. Most often, argument occurs in dialogue. When this happens, the reasoning in the argument can be called dialectical reasoning. Some say that all reasoning

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

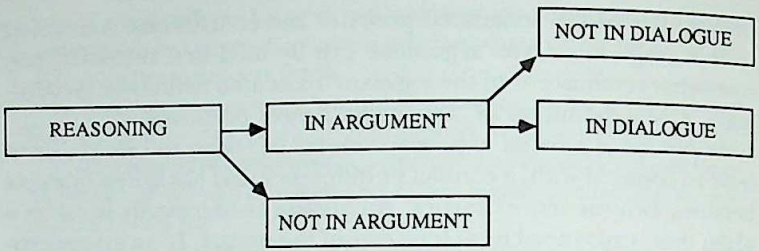


Figure 1: Containment relations

is dialectical, but it is more widely acceptable to say that some reasoning can be nondialectical, meaning that it can be a solitary rather than an interactive process. (See Figure 1.) The usual examples of nondialectical reasoning cited are activities like planning or problem solving. Even though they are often interactive activities, these contexts of reasoning can also (presumably) be solitary in some cases. Some think, however, that actions like planning or solving a problem on your own are really disguised forms of interactively reasoning with yourself, where the same person alternately plays the role of proponent and respondent. There is something in this approach, but it will not be argued for here.

At any rate, argument often occurs in dialogue, and to understand an argument, it is very often highly important to know something about the context of dialogue in which the argument has occurred.

There are many different kinds of dialogues.[21] Some kinds of dialogue are nonargumentative, but many are inherently argumentative in nature, some more so than others.[22] The critical discussion is inherently argumentative. It is a species of persuasion dialogue where each participant tries to show, by means of arguments directed to the other participants, that his (the first participant's) point of view is right.

The negotiation dialogue is argumentative in a different kind of way. In contrast to persuasion dialogue, negotiation is a form of interest-based bargaining where the goal is to "get the best deal." This type of dialogue is characterized by Christopher W. Moore[23] as

[21] See William C. Mann, "Dialogue Games: Conventions of Human Interaction," *Argumentation*, II (1988): 511–532, for a classification of types of dialogue. See also the classification given in my *Informal Logic: A Handbook for Critical Argumentation* (New York: Cambridge, 1989).

[22] A number of types of argumentative dialogue are being classified and studied in a current research project undertaken with Erik Krabbe at NIAS, to be published as a monograph, *Commitment in Dialogue*. The classification Krabbe and I have developed is somewhat different from that given in Table 1 below, but contains many of the same types of dialogue.

[23] *The Mediation Process: Practical Strategies for Resolving Conflict* (San Francisco: Jossey-Bass, 1986).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

a kind of competitive argumentation where the arguers make concessions in order to try to maximize their own share of a set of goods that is not sufficient for all. The goal here is not to show that a proposition is true or right, based on evidence, as it is in a critical discussion.

The debate may initially appear to be a kind of critical discussion. While it does have some elements of the critical discussions, its objectives and rules are generally quite different. There are two primary participants involved in a debate, but their real goal is to persuade a third party (usually an audience or a judge). The debate can be rule-governed, but the rules are characteristically permissive, from a logical point of view, and allow for highly aggressive adversarial clashes. The way the audience or judge decides the outcome may, in typical cases, not reflect a high degree of rationality. A debate often takes place in an institutionalized context, for example, a congress or parliament. But the institutional rules of debate often tend

Table 1: Eight Types of Argumentative Dialogue

Type of Dialogue	Initial Situation	Goal	Benefits
1. Critical Discussion	Difference of Opinion	To Convince Other Party	Understand Positions Better
2. Debate	Adversarial Contest	Persuade Third Party	Clarification of Issue
3. Inquiry	Lacking Proof	Prove or Disprove Conjecture	Knowledge
4. Negotiation	Conflict of Interest	Maximize Gains	Settlement and Consensus
5. Planning Committee	Collective Action Required	Joint Plan or Decision	Airing of Objections
6. Pedagogical	Ignorance of One Party	Teaching and Learning	Spread of Knowledge
7. Quarrel	Personal Conflict	Hit Out Verbally	Venting of Emotions
8. Expert Consultation	Need for Expert Advice	Decision For Action	Second-hand Knowledge

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

to be very permissive in allowing emotional argumentation of questionable relevance.

The inquiry is an essentially *cumulative* type of dialogue, meaning that retraction of commitment is not generally permitted. The inquiry is a collaborative investigation that seeks to prove something, or alternatively to show that the existing evidence is insufficient to prove it. The inquiry is a hierarchical procedure of reasoning (similar to an Aristotelian demonstration) where the premises are supposed to be better known or established than the conclusion to be proved from them. By contrast, the persuasion dialogue has the less ambitious goal of trying to show that one opinion is more plausible than another on a subject of controversy or conflict of opinions. The inquiry tries to show whether the proposition in question should be considered part of established knowledge or not.

The planning committee type of dialogue resolves its differences of points of view by a discussion that ends in a vote, or similar way of proceeding. Although this type of dialogue often involves logical reasoning and persuasive debate, the outcome may simply be based on self-interest. Ideally, however, the conclusion reached through this type of dialogue should use practical reasoning by reaching agreement on common goals and on the best means to implement them.

In the planning committee type of dialogue, two or more participants have one or more goals they are trying to carry out in a particular situation, and they are considering alternative ways of realizing their goals. The problem is to mesh the goals with the known available means of carrying them out, relative to the given situation.

The pedagogical type of dialogue presumes that the teacher has access to knowledge, and that his role is to impart that knowledge to the student. This kind of dialogue shares some characteristic with the expert consultation. But in the latter type of dialogue, a layperson consults an expert in a skill or domain of knowledge in order to obtain advice on a decision or a problem. In the expert consultation, both parties have obligations. The layperson must ask clear and relevant questions. The expert must give clear and nontechnical replies that guide the layperson to an understanding of the subject needed to solve his problem.

The quarrel is a kind of argumentative combat where each participant tries to hit out verbally at the other. The quarrel is characterized by an almost total absence of logical reasoning and by heightened emotions.

VI. ARISTOTLE ON REASONING

As noted, the logicians' definition of argument usually found in logic textbooks originally sprang out of Aristotle's theory of the syllogism.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

There is, however, another, much broader and more heterogeneous conception of argument (and reasoning) in Aristotle's *Topics, Rhetoric,* and *De Sophisticis Elenchis.* This conception has a lot in common with the framework proposed above.

Aristotle frequently uses 'reasoning' where we would use 'argument', and vice versa. For example, he writes about different kinds of reasoning where, in the above framework, it would be, strictly speaking, more appropriate to talk about different kinds of argument. But these apparent differences can be explained, given the presumption that, since reasoning occurs in argument, different kinds of argument context will result (derivatively) in different kinds of reasoning.

In the *Topics* (100a25), Aristotle defined reasoning (*syllogismos*) as "an argument (*logos*) in which, certain things being laid down, something other than these necessarily comes about through them." Aristotle identified four kinds or special contexts of reasoning, identified by the four different kinds of premises they have. Reasoning is a *demonstration* (*apodeixis*) when the premises are "true" and "primary" propositions established in a scientific discipline (100a28). Reasoning is *dialectical* (*dialectikos*) when it starts from opinions that are generally accepted (100a31). Reasoning is *contentious* (*eristikos*) when it starts from opinions that appear to be generally accepted, but are not (100b24). Reasoning is *misreasoning* (*paralogismoi*) when it starts from premises in a special science that are neither true nor primary (101a10). Aristotle's conception of reasoning clearly placed key importance on where the premises come from. For Aristotle, premises are not merely arbitrary or "designated" assumptions in reasoning.

Aristotle's definition of reasoning is narrow in two respects. By using the word 'necessarily', he restricts himself to deductive reasoning, ruling out the possibility of inductive or plausible reasoning. By using the phrase 'something other than', he restricts himself to noncircular reasoning, ruling out as reasoning the case where what "comes about" is not "other than" what is "laid down."

Dialectical reasoning is opinion-based reasoning in a context of argumentative dialogue on a controversial issue. Demonstration appears to correspond to how we in the twentieth century think of scientific investigation or inquiry. Or, at any rate, it corresponds to how the results of scientific investigation are presented and organized as established findings. Demonstration starts only from premises that can be established as "verified" or "known to be true" within a field of scientific knowledge.

Contentious reasoning or *eristic* is the degeneration of reasoned dialectic into verbal combat, or the purely adversarial quarrel. Curi-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ously, Aristotle also thought that demonstrative reasoning had a negative counterpart, misreasoning. What might he have meant by this? Perhaps it could correspond to the kind of pseudoscientific reasoning typified by the cases studied in Martin Gardner's *Fads and Fallacies in the Name of Science.*[24] At any rate, it appears to be highly significant to the study of informal fallacies generally that there might be different kinds or contexts of a fallacy, depending on whether the context of argument is that of demonstrative or dialectical reasoning.

Dilip K. Basu[25] thinks that the Aristotelian demonstration was not intended as a model of scientific investigation, but as a model of how a teacher should represent scientific knowledge to a student. According to this interpretation, the scientific tutor gets his premises from the field scientist, and then uses them to demonstrate conclusions in an orderly way. This fits Euclidean geometry (the paradigm of demonstration) very well. For the numbering of theorems in Euclidean geometry does not represent the order of discovery, but the tidied-up order of presentation of the proofs in a carefully organized sequence.

Galen (*circa* 129–199) wrote a major work on the demonstration called *De Demonstratione* (*Peri Apodeixeōs*), unfortunately lost some time after the sixth century. But in a short treatise, "On the Best Teaching," he argued for a "linear" or "geometrical" exposition of scientific knowledge in teaching, instead of freely allowing students to argue for both sides of a disputed issue. According to John Spangler Kieffer's introduction to Galen's *Institutio Logica,*[26] Galen thought that demonstration had a double role. Not only is demonstration useful for establishing scientific facts, but it can also be useful to refute errors and to defend science against the adherents of erroneous theories.

It seems, then, that the Aristotelian concept of a demonstration could be somewhat complex and ambiguous. It could refer to a context of reasoning as a kind of group investigation by a collection of peers who have the goal of proving or disproving some conjecture. This kind of inquiry would be typical of scientific research, but it need not be confined to purely scientific contexts. Other kinds of official or nonofficial inquiries could take place as well. But, second, the demonstration could also refer to the presentation of the results of an inquiry in the forms of teaching, published results, writing a report, etc. This might not be argument among peers, but could

[24] New York: Dover, 1965.
[25] "A Question of Begging," *Informal Logic,* VIII (1986): 19–26.
[26] Baltimore: Johns Hopkins, 1964.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

rather be a form of advice given by experts (or those who have been privy to the inquiry) to those who are being instructed in the findings of the inquiry.

Thus, the apparent ambiguity found in the Aristotelian concept of the demonstration can be explained as a kind of shift from one context of dialogue to another. In one sense, the demonstration is simply an inquiry. But through dialectical shifts, 'demonstration' could refer to other more complex contexts of argument, where the results of an inquiry are presented to a new audience. This could take various forms of dialogue. One could be a context of advice-giving dialogue by expert consultation. Another could be the pedagogical context of teaching the results of the inquiry to students.

Without trying to interpret definitively what Aristotle really meant to say—a task best left to the specialists in Greek philosophy—it is interesting to see that there was an Aristotelian tradition of understanding argument and reasoning which is not that dissimilar in broad outline from the apparently radical framework proposed above. While our new definitions of these old terms may certainly seem radical from a viewpoint of established twentieth-century preconceptions in logic, they do not seem at all out of place in relation to general perspectives on argument and reasoning in ancient writings on argumentation, and notably in Aristotle.

VII. WHAT IS LOGIC?

It is a platitude that logic is the science of reasoning. But as a definition, this platitude is inadequate. For psychology is also a science of reasoning. As a definition, it is much better to say: *logic* is the study of how normative models of reasoning are used in different contexts of dialogue. The chief objective of logic should be to evaluate arguments in order to separate the good (strong, correct, valid) arguments from the bad (weak, erroneous, fallacious) arguments. To put this in the form of a new platitude, logic is the evaluation of reasoning in arguments.

The key to this new approach is that logic uses normative models of reasoning that is good for purposes of various kinds of arguments. From the point of view of logic, the science of reasoning should be the study of whether or not conclusions can be extracted correctly from premises (assumptions), once it is clear what these assumptions are in a context of argument. Logic has to do with defining the assumptions as well, or identifying them correctly, and with evaluating whether a putative conclusion can be correctly derived from a given set of premises.

As already noted, logical pragmatics is the study of the uses of logical reasoning in a context of argument.

Formal logic abstracts from the content of the premises and con-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

 Digitized by Arya Samaj Foundation Chennai and eGangotri

clusion of an argument, calling them propositions. *Informal logic* must interpret the uses of these propositions as speech acts in a context of dialogue, seeing them as moves that incur or relinquish commitments, e.g., assertions, denials, retractions, questioning moves, etc.

Formal logic has to do with forms of argument (syntax) and truth values (semantics). At any rate, this is the traditional conception. Informal logic (or more broadly, argumentation, as a field) has to do with the uses of argumentation schemes in a context of dialogue, an essentially pragmatic undertaking.

But, as has so often appeared paradoxical, informal logic does have to do with abstract, normative models of reasonable argumentative dialogue, which do have a (dialectical) formal structure. Hence informal logic is partly formal! It has to do with formal games of dialogue. But it is also practical, i.e., nonformal, in some respects. The key to this apparent paradox is to be found in the distinction clearly drawn by Hamblin between the formal and the descriptive study of dialogues (*op. cit.*, p. 256).

Hence the strongly opposed current distinction between informal and formal logic is really an illusion, to a great extent. It is better to distinguish between the syntactic/semantic study of reasoning, on the one hand, and the pragmatic study of reasoning in arguments, on the other. The two studies, if they are to be useful to serve the primary goal of logic, should be regarded as inherently interdependent, and not opposed, as the current conventional wisdom seems to have it.

In the twentieth century, the pragmatics of the uses of arguments in dialogue has not been widely held to be of any serious importance, either as a useful tool of reasoning or as an object of serious or vigorous academic study. Although Aristotle founded the study of dialectical reasoning quite firmly in the *Topics* and *De Sophisticis Elenchis*, this area of study has never again been taken up seriously in a sustained manner. It has always remained at the fringes of logic, in the study of informal fallacies, in an unexplained and often primitive state. Only very recently has this long-dominant trend begun to reverse itself to some extent.

Logic in the Western world has long been dominated by Aristotle's syllogistic logic and, since the beginning of the twentieth century, by symbolic (mathematical) logic. In using these formal structures of deductive logic, it has been presumed that formal logic is the most important kind of logic, or perhaps the only important kind of logic, and that formal logic is the abstract study of certain kinds of relations—primarily, the deducibility relation characteristic of the de-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ductively valid type of argument. This conception of logic takes the point of view that, in the study of reasoning in logic, it does not matter where the premises come from—they are arbitrarily designated (given) propositions. The only thing that matters is the relation between the set of propositions called the premises, and the single proposition designated as the conclusion—and, in particular, the relationship between the truth values of these pairs. Reasoning has now become exclusively concerned with the question of how the conclusion necessarily comes about from the premises. It is no longer about how the premises were originally laid down.

This one-sided view of reasoning has had some peculiar results. Given the dominance of this view, it is no longer possible to see why or how arguing in a circle could be a fallacy. It is never possible to go from truth to falsehood if your conclusion is the very same proposition as your premise. True, you have not gone anywhere, and such a premise would seem useless as a starting point for some real reasoning.

Only recently has it become more apparent that a pragmatic approach is absolutely necessary in order to make sense of informal fallacies. What are fallacies? They are violations of rules of reasonable dialogue.[27] But over and above this, they are also deceptive tactics used unfairly in argument to defeat an adversary in dialogue. If the study of fallacies is to be part of logic, clearly logic can make no headway in working toward its primary goal unless the pragmatic study of the uses of reasoning in argument (informal logic) is included as a legitimate part of the subject.

Thus, in redefining logic, we need to go back to the Aristotelian roots of the subject, to expand the boundaries of the subject to include informal logic, and to give up some of the false oppositions between the informal and formal logic special-interest groups.

DOUGLAS N. WALTON

Netherlands Institute for Advanced Study
& University of Winnipeg

[27] See van Eemeren and Grootendorst, *op. cit.*, p. 177; and my *Logical Dialogue-Games and Fallacies* (Lanham, MD: UP of America, 1984), p. 237.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

COMMENTS AND CRITICISM

VIRTUE AND IGNORANCE

IN "The Virtues of Ignorance," Julia Driver[1] discusses the important question of whether there exists some class of virtues that require as a necessary ingredient that she who possesses or displays any one of them lacks knowledge. Driver thinks there is such a class. She calls the class of virtues that involve ignorance in an essential way *virtues of ignorance*. The argument is important, if true. Many recent writers have pointed to an ineliminable tension between our ethical and personal ideals. If successful, Driver's argument would establish the existence of a different, but equally ineliminable tension: between our ethical and epistemic ideals. The argument for virtues of ignorance, however, goes wrong in several important places. Let me explain why.

I

Driver claims first that it is a common assumption within virtue ethics, including on Aristotle's account, that

> (A′) the agent who possesses some virtue v "exercises it knowingly" (373).

(A′) is provided with a variety of interpretations. The one that Driver makes central use of is implausibly strong. We are told that (A′) is true because (a) the virtuous person acts with practical wisdom; (b) the virtuous person chooses the mean between excess and defect; (c) he performs the act that expresses the mean disposition; and (d) "in order for a person to act according to virtue, that person must know what he is doing under the description that makes the action characteristic of the virtue" (373). (a)–(d) are treated by Driver as if they were more or less equivalent. But this is not true.

(A′) is a common assumption of virtue theories if the requirement of exercising virtue knowingly is interpreted as requiring a kind of *know-how*. The agent who possesses some virtue v must possess some sort of procedural knowledge and the relevant sort of knowledge can be cashed out perhaps in terms of (a)–(c). That is, the person who possesses some virtue v must have certain habits of perception, feeling, and action; she must be a good detector of the moral saliencies relevant to that virtue, of the mean action, and so on. But it is not a general requirement that to be ascribed some virtue v the agent must

[1] This JOURNAL, LXXXVI, 7 (July 1989): 373–384.

0022-362X/90/8708/420–8
© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

possess declarative or propositional knowledge. In particular, it is not a consequence of (A′), properly interpreted, that (d) "*in order for a person to act according to virtue, that person must know what he is doing under the description that makes the action characteristic of the virtue*" (373, my italics). (d) requires a sort of linguistic sophistication, reflectiveness, taxonomic ability, and conscious awareness which virtue theorists have historically expressed is decidedly not necessary for virtue.

The first mistake then involves using (d) as the proper interpretation of (A′). The second mistake follows immediately when it is claimed that a "consequence of such a view of virtue is:

(A) No virtue is constituted by, or based upon ignorance" (373).

Initially, Driver unpacks (A) in terms compatible with a procedural, or know-how analysis. So the generous person must, in order to be properly ascribed the virtue of generosity, "pick up" on the needs of others and give aid. She writes: "(A) is necessary to most accounts of virtue, because on these accounts being virtuous is largely a matter of *knowing what to do*" (373, my italics). This way of putting things hardly entails that the generous agent needs to know what "he is doing under the description that makes the action" an act of generosity. But this is the interpretation Driver goes on to deploy. For the class of virtues she calls virtues of ignorance are virtues on her view because it is necessarily the case that the person who possesses or displays a virtue of ignorance does not comprehend it under the appropriate type description.[2] Driver sees the existence of virtues of ignorance as falsifying (A). It looks initially as if the argument is a very simple case of modus tollens:

1. If "(A) No virtue is constituted by, or based upon ignorance" (373), then there are no virtues of ignorance.
2. There are virtues of ignorance.

---

3. Therefore, some virtues are constituted by, or based upon ignorance.

But actually it is a more complex case of the argument form:

1. If (A′) the agent who possesses some virtue v "exercises it knowingly," then (d) "that person must know what he is doing under the description that makes the action characteristic of the virtue" (373).
2. If (d) "in order for a person to act according to virtue, that person must know what he is doing under the description that makes the

[2] Strictly speaking, there is the superordinate-type virtue: of ignorance. This type consists of the subordinate virtue of ignorance types; for example, modesty, blind charity, the refusal to hold a grudge, and so on. Finally, there are the tokens of these subordinate types, e.g., particularly instances of modesty or blind charity.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

action characteristic of the virtue" (373), then (A) "No virtue is constituted by, or based upon ignorance" (373).

3. If "(A) No virtue is constituted by, or based upon ignorance" (373), then there are no virtues of ignorance.
4. There are virtues of ignorance.

---

5. Therefore, it is false that "(A) No virtue is constituted by, or based upon ignorance" (373); and it is false that (d) "in order for a person to act according to virtue, that person must know what he is doing under the description that makes the action characteristic of the virtue" (373); and it is false that (A′) the agent who possesses some virtue v "exercises it knowingly" (373).

The point of emphasizing the argument's complexity is to make clear that the existence of a class of virtues of ignorance would not falsify either (A′) or (A) so long as these are not interpreted in accordance with (d). No plausible form of virtue theory does so, so we should reject Driver's initial characterization. This defeats her claim that the existence of the virtues of ignorance, if it could be established, would show that virtue theory is deeply mistaken in its commitment to (A′) or (A).

II

This much seems firm. But the rejection of the interpretation of (A′) and (A) in terms of (d) is compatible with the existence of a class of virtues of ignorance. Most everyone thinks that the virtues are heterogeneous in kind, so it is possible that Driver is correct that there is a type of virtue that is correctly characterized as essentially involving an epistemic deficit.

The person who possesses any virtue must, according to the analysis I have recommended so far, display (or be disposed to display) that virtue in appropriate situations, and thus she must possess whatever know-how such expression requires. But it is compatible with this that there is a class of virtues whose proper ascription requires that the agent who possesses a member of this class must be ignorant of certain things, possibly even that she possesses that very virtue. The generous person might not know that what she does falls under the description of generosity. But it would not disqualify an ascription of generosity if she did. Virtues of ignorance, however, are allegedly different in this respect. If the agent knows that she has the virtue, she does not; or so Driver argues. This is an interesting and important claim in its own right, so it is worth a closer look.

Driver carefully works out an argument for a paradigm-case virtue of ignorance: modesty. The sort of modesty she has in mind is not sexual modesty, but the sort of modesty "associated with self-deprecation, or an underestimation of one's self-worth" (374). The first question is: What is it of which the modest person must be ignorant?

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Sometimes, indeed usually, Driver claims that it is essential for a correct ascription of modesty that the modest person not know that she possesses the virtue of modesty itself. The modest person is not allowed, unlike the generous person, to know that she is modest. The second interpretation is that the modest person must be (generally) ignorant of, and thereby underestimate her worth, accomplishments, and so on. Call the first interpretation the *strong-ignorance claim*. It requires ignorance at the second level, at the level of the general virtue description that applies. Call the second interpretation, the *weak-ignorance claim*. It requires ignorance of certain facts or features of the self, others, or the world.

One might think that the modest person could satisfy the weak-ignorance claim without satisfying the strong-ignorance claim. That is, some person P might know that she typically underestimates her worth and accomplishments and thus know that she is modest. But Driver thinks this is not possible. She correctly points out that the statement 'I am modest' is odd from a performative point of view. But it is important to emphasize that it is not nearly a contradiction, nor is it nearly as odd as G. E. Moore's paradox ("It's raining, but I don't believe it"). And it is not odd at all in certain contexts. For example, in a communicative situation with close friends or a therapist, a person might be brought to see that she is modest, and thus come to understand some of her behavior under that description. Such self-comprehension might even result from self-reflection in a savvy, strong evaluator in Charles Taylor's sense. It would be extremely odd, counterintuitive, and moralistic if it disqualified a person from being modest if she knew she was. To be sure, if she started to brag, "I am modest," she would generate her own disqualification. But this is for other reasons than that she possesses the reflexive knowledge that she possesses the virtue in question. I conclude that we have been given no reason to think that believing that one possesses some virtue v is ever sufficient by itself to disqualify the ascription of v. This undermines one interesting, but excessively strong claim about the virtues of ignorance—what I have called the strong-ignorance claim. Driver is wrong to say that "I can be modest, but I cannot *know* it" (376). And she is wrong to claim that there is a self-other asymmetry in the case of the virtues of ignorance which takes this form: "I can ascribe the virtue to another, but I cannot coherently, and sincerely ascribe it to myself" (380).[3]

[3] It is a general feature of virtue ascription—indeed, according to attribution theorists, it is a general feature of trait ascription, period—that we are less disposed to ascribe general traits to ourselves than to others. But this is an altogether different point.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

If it is not the virtue itself of which the person who possesses a virtue of ignorance must be ignorant, what is it of which she must be ignorant? Perhaps the weak-ignorance claim is true, even if the strong-ignorance claim is false.

III

Driver favors what she calls an "underestimation" account of modesty over either a behavioral or understatement account on grounds that a person can behave modestly without being modest, and a person can understate her accomplishments knowingly and manipulatively. According to the underestimation account, the modest person:

> If he speaks, then he underestimates the truth, but he does so unknowingly. This entails that the modest person is ignorant, to a certain degree, with regard to his own self-worth. He underrates himself, and therefore only takes part of the credit due him (376).

> In order for a person to be modest, she must be ignorant with respect to her own self-worth. She must think herself less deserving, or less worthy, than she actually is (377).

I have already argued that the modest person can know that she is modest. And although I think the underestimation account is wrong, this point is compatible with it. A person might know that she is modest in Driver's sense that she possesses "a dogmatic disposition to underestimation of self-worth" (378). That is, an individual might know that she possesses such a disposition, but be unaware in specific occurrent cases that she is engaging in underestimation or, even if she is aware that she is probably underestimating herself in some particular situation, she might be in the dark as to what exactly it is about her worth, accomplishments, and so on, which she is underestimating.

Even though the underestimation account is, or can be made, compatible with my claim that modesty can involve knowledge that one is modest, there is something wrong, I think, with the underestimation account. Compare it to a fourth account, which Driver never considers, and which I call the *nonoverestimation account.*

According to the nonoverestimation account, the modest person may well have a perfectly accurate sense of her accomplishments and worth but she does not overestimate them. This account is better than the behavioral account, since nonoverestimation is more than behavioral. It involves the manner in which the modest person sees herself and experiences her worth and accomplishments. And it is better than the understatement account for the same reason: the state of mind presupposed rules out manipulative reserve in one's publicly expressed self-estimations.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Finally, it is better than Driver's underestimation account for four reasons. First, there is in the end something deeply problematic about a characterization of the virtue of modesty in terms of a "dogmatic disposition to underestimation of self-worth" (378). On Driver's view, there are no modest people who do not possess this dogmatic disposition. But this, it seems to me, is deeply counterintuitive. Perhaps there is a kind of modest person who fits her description. But I doubt it. I think we would be inclined to think that the epistemic deficit of systematic underestimation made such cases into cases of "false modesty." The truly modest person cannot be so systematically in the dark about her worth. In fact, in the sorts of cases Driver describes, I suspect that the question of modesty would not even enter into evaluative discussion. I think we would just say of her so-called "modest" person that she lacked appropriate self-esteem or self-respect.

In the second place, and relatedly, on Driver's view, the person who "has an accurate assessment of his accomplishments and character" can behave modestly, but cannot be modest. This, too, is paradoxical. Its only warrant, as far as I can see, is the strong-ignorance claim rejected above, which disqualifies ascription of virtues such as modesty to all persons who know they have it. The nonoverestimation account avoids this paradox. Modesty is compatible with accuracy of self-assessment. The modest person simply does not overestimate her worth.

It is easy to imagine that the world's fastest runner might be modest and know that he is (suppose that sports commentators are forever pointing it out to him because they are exasperated by his modesty). He is not ignorant either of his modesty or of his accomplishments. Perhaps he not only has no dispositions to immodesty, but also assesses his accomplishments as less significant than others do. He might think that being the world's fastest human is not so important *sub specie aeternitatis,* or he might think that his being the world's fastest human involves a significant amount of luck. It is easy to imagine that a person who thought this way might eventually come to think that he was not really modest at all, but rather that he appraises things realistically. He simply does not overestimate his accomplishments, as others do. This suggests that over- or underestimation involves a significant amount of observer relativity.[4]

Third, Driver's underestimation account has the further paradoxical consequence that, "[s]ince modesty necessarily involves ignorance, it is also necessarily involuntary in nature" (381). Indeed,

[4] Discussion of this paper with Amélie O. Rorty, David Wong, and Ken Winkler was helpful on this point.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Driver is attracted to the view that only modesty that arises spontaneously, from a person's temperament, is true modesty. This is because working at being modest would require that one "would need to practice self-deception" (382). Driver is led to this conclusion because of two mistakes. There is the mistake of thinking that the cultivation of modesty would need to aim at the cultivation of the "dogmatic disposition to underestimation of self-worth" (378). For the person who is not now modest, this would require practice and eventual success in the acquisition of a vice, the vice of not seeing the self accurately. Apparently, it is Driver's view that, if one is naturally disposed to dogmatic self-delusion, this has no bearing on our ascriptions of true modesty. Whereas if this delusion is the product of work on one's character, it is a total disqualifier. The second mistake arises from thinking that learning to underestimate self-worth in the way required would have to eventuate, in its last step, with the person's somehow making herself ignorant of the fact that she had made herself modest. Otherwise, she would be modest occurrently, so to speak, in all episodes requiring underestimation, but she would know the general truth that she was modest. This is not permitted according to the requirement of strong ignorance. To me, the twin ideas that the virtues of ignorance are essentially characterized by ignorance at both levels, as well as closed, in principle, to self-cultivation, are extremely implausible consequences of the underestimation account. Indeed, I take them to be a reduction of the account.

The nonoverestimation account avoids all such trouble. Working at becoming modest is possible because one does not need to acquire any dogmatic disposition of underestimation. One simply needs to learn not to overestimate one's accomplishment and worth. Furthermore, there is no deep paradox in knowing that one is modest on the nonoverestimation account. The modest person, of course, does not, as Bernard Williams says, act "under the title of modesty." She does not overestimate simply because she wants to be modest. But that is a different issue.

The fourth point in favor of the nonoverestimation account is that it fits nicely with findings in cognitive social psychology regarding self-serving attributional biases. In a recent paper, Shelley Taylor and Jonathan Brown[5] provide evidence, based on a meta-analysis of a large array of experiments, that most people dramatically overesti-

[5] "Illusion and Well-Being: A Social Psychological Perspective on Mental Health," *Psychological Bulletin*, CIII, 2 (February 1988): 193–210. I discuss this research at greater length in *Varieties of Moral Personality: Ethics and Psychological Realism* (Cambridge: Harvard, forthcoming).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

mate themselves across a wide variety of situations. The evidence suggests that most people are prone to unrealistically positive self-evaluations. Here are some of the relevant findings: given a list of trait names, subjects judge positive traits to be overwhelmingly more characteristic of self and intimates than negative traits; subjects rate self and self's performance more positively than observers do; persons score themselves and their intimates better than all others on all measures; they judge the group(s) to which they belong as better than other groups; they have more trouble recalling failure than success; recollection of task performance is exaggerated in recollection; favored abilities are seen as rare, disabilities are seen as common; favored abilities are seen as important, disabilities as unimportant; improvement is exaggerated on tasks that are considered important; initially modest attributions of success or failure become more self-serving over time, e.g., on a joint performance credit given to one's partner shifts to oneself over time.

The strength and pervasiveness of these self-serving attributional biases support the intuition that what we need is not more people who underestimate their self-worth, but more people who do not overestimate their worth and accomplishments. Most of us are immodest overestimators. Modesty is the best policy, but it need not involve, ideally it should not involve, the relatively rare defect of underestimation. There is no need for the modest person to be ignorant of her worth and accomplishments. The modest person simply sees more clearly than most how the base rates work; she understands, perhaps, and thereby avoids, certain natural tendencies to frame things in self-serving ways, and she does not need to perceive her self-worth as so much better than average. Modesty and realism in self-appraisal are perfectly compatible.

IV

Driver gives two other examples of virtues of ignorance: blind charity, charity that involves seeing the good in others but not the bad; and the refusal to hold a grudge. Neither of these fits the desired characterization of virtues of ignorance. But they fail to do so for somewhat different reasons. One common feature is that neither remotely requires strong ignorance. It is easy to imagine persons who possess the relevant dispositions, to overlook the bad in people and to forgive and forget quickly, respectively, and who know that they do. Furthermore, with respect to blind charity, it is hardly uncontroversial as a virtue. Charity of thought, characterized essentially in terms of an epistemic defect of never seeing the bad in others, might well lead the blindly charitable person to support and sustain all sorts of bad habits in others. Even if this did not happen,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

we would think of such a person as sweet and dear perhaps, but also as naive and blindsighted.

The third example, that of refusal to hold a grudge, is supposed to involve a person with a general disposition to forgive and forget. It is often good to forgive and forget. If the disposition to do so is indiscriminate, however, it, like the disposition to blind charity, will be taken rightly to involve a defect—to be applied without practical wisdom or without a mean disposition. Furthermore, it is hard to see, even in a best-case scenario, how this disposition would really count as a virtue of ignorance. The person who forgives and forgets hardly *really* forgets in the epistemic sense that she does not remember, or cannot remember, the insult or injury. She does not become amnesiac with respect to the relevant hurt. It is just that she is able to put the insult or injury out of mind on an ongoing occurrent basis. In this way, the memory is kept from poisoning ongoing relations.

V

There are ignorant people who are virtuous.[6] Tolstoy has written about them. But nothing in the argument of "The Virtues of Ignorance" convinces me that there is a class of virtues that require essentially a kind of ignorance. In each of Driver's cases, there are grounds for doubting that the alleged virtue is a *bona fide* one. But even if one accepts that she has isolated a virtue in each case, there is no essential link between possession of any of her alleged virtues and ignorance that one possesses the virtue itself (the strong-ignorance claim), or of ignorance of some particular feature of the self or world (the weak-ignorance claim).

OWEN FLANAGAN

Wellesley College

[6] The relation between reflectiveness and goodness is discussed in my "Identity and Strong and Weak Evaluation," in Owen Flanagan and Amélie Oksenberg Rorty, eds., *Identity, Character, and Morality: Essays in Moral Psychology* (Cambridge: MIT, 1990).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

BOOK REVIEWS

The Mind of God and the Works of Man. EDWARD CRAIG. New York: Oxford University Press, 1987. x+ 353 p. Cloth $49.00.

This is a rich and fascinating book, covering a host of topics from the seventeenth century to the present, with the discussion ranging from general overviews to detailed discussions of particular philosophers or issues. Its ambit is so broad that it is difficult to describe its theme succinctly, but a core thesis can be discerned. This is a thesis about the development of views of human understanding since the early seventeenth century. The author looks at three "stages" in this development. The dominant tendency in the seventeenth and early eighteenth centuries was to model human understanding on the understanding of a transcendent God, and knowledge was conceived of as a kind of insight, the knower being construed as a spectator (ch. 1). This tendency was, Edward Craig argues (ch. 2), ultimately undermined by Hume. In the second stage, roughly between 1780 and 1820, the idea of a transcendent God was replaced by a much more pantheistic notion of an immanent God, and the knower came to be conceived of as an active participant in the act of understanding (ch. 3). This development, in which Craig sees Hölderin playing a major role, finds its fullest expression in Hegel (ch. 4). Finally, in the modern era, this conception of knowledge in which the knower is active is developed in a more radical way, so that even such areas as logic and mathematics, which were formerly paradigms of the idea that understanding consists in insight, came to be treated in a constructivist fashion (chs. 5 & 6). Indeed, the general tendency is toward an attempt to elucidate human understanding by reference to the human activities which constitute or manifest, or which are in some other way an integral part of, understanding.

In undertaking such an ambitious program as this, there are numerous dangers. One is that of teleology, of seeing the history of philosophy as leading inevitably to the present, so that only presently recognized concerns and problems can be treated as the legitimate ones. Such an approach, apart from its obvious historiographic drawbacks, effectively prevents us from standing at any critical distance from modern concerns. The converse danger is to see the development of philosophy since the sixteenth century in terms of a program of secularization, to see the origins of modern concerns with knowledge as lying in a theological context, a context which

0022-362X/90/8708/429–432
© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

must be referred back to if one is to have any complete understanding of these concerns. Craig succombs to neither of these dangers. Indeed, he is exceptionally clear headed about what kinds of issue have to be raised in assessing philosophical change, and what kind of enlightenment can be gained from studying it. In discussing the emergence of causal accounts of knowledge and meaning, for example, he is able to show how the demise of the view that understanding is paradigmatically a form of insight made possible "various approaches to philosophical problems which, in the earlier climate, would at once have been felt to be so wholly unsatisfactory as not to be worth pursuing" (284). This is surely the right approach to such questions and it is pursued in a sensitive and illuminating fashion.

In broad outline, Craig's general thesis about the shift from a conception of understanding modeled on divine insight to a more constructivist conception is convincing. The details of his account are occasionally problematic, however. I shall restrict my attention to the early parts of the book. In the first chapter, Craig provides a general outline of the "Image of God" thesis in the seventeenth and early eighteenth centuries. This thesis turns on the idea that, being created by God, we are similar to him in certain crucial respects, and it is our reason or intellect that is usually picked out as being where the core similarity lies. Craig focuses on a view to be found in a number of seventeenth-century natural philosophers from Kepler and Galileo onward that our grasp of mathematical truths is the same as that of God: that, in the case of mathematics, we have the same kind of grasp as God, only to a very reduced degree, since God knows all mathematical truths whereas we only know a few. He realizes (22 ff) that Descartes is a problem in this respect, because of his doctrine that God creates eternal truths, but he points out that in his doctrine of free will Descartes does come close to the *similarity thesis,* i.e., the thesis that we are like God. I do not find this a particularly plausible accommodation of Descartes's doctrines to the similarity thesis. Descartes's doctrine of eternal truths tells us that we can have no insight into how God knows such truths to be true, and that consequently we cannot say that such truths are the same for God as they are for us. The upshot of Descartes's position is such a radical distancing of God from anything we can say about our knowledge and reasoning processes that God's knowledge and reasoning processes effectively become irrelevant in any account we might give of ours. In other words, right at the beginning of the period at which Craig believes the similarity thesis comes to the fore, we find what is tantamount to a denial that human knowledge can be modeled on divine knowledge.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The case of Leibniz is different but equally problematic. Leibniz does accept the similarity thesis, but he appears to reverse the direction of similarity. Leibniz's position is in a crucial respect the opposite of that of Descartes: whereas Descartes argues that we cannot know how God knows eternal truths to be true, and that therefore they cannot be the same for us as for God (the doctrine of equivocality), Leibniz takes it as given that eternal truths must be the same for us as for God (the doctrine of univocity) and reasons that we therefore know how God knows them to be true. He argues that God knows them to be true because he knows their proofs, something directly contrary to the idea that knowledge consists in some spectator-like insight, and something which effectively models God's reasoning process on ours rather than vice versa.

It is central to the detail of Craig's argument that the similarity thesis, the spectator view, and the knowledge-as-insight conception go together. I am not convinced by this. In Descartes's case, one has an insight and spectator view of knowledge without the similarity thesis; in Leibniz's, one has a reversal of the similarity thesis. One can also find in this period a prevalent theory of knowledge which, while retaining the similarity thesis, denies the spectator view and radically modifies the knowledge-as-insight conception. Craig ignores this tradition, which might be termed the *makers' knowledge* view. It is committed to the idea that making something puts one in a special cognitive relation with what one has made. Hobbes argued that geometry and civil society were areas where demonstration is possible because—unlike natural philosophy, which can only be conjectural—human beings have created them. Locke revised and generalized this argument in the case of civil society by arguing that the special knowledge derives not from the actual making or doing but from knowing the archetype in accordance with which what is made or done is made or done, and in the case of civil society all of us, and not just its founders, have access to such knowledge. This principle was subsequently developed fully by Vico, who maintained that history and society are human artifacts and that consequently we can have a special knowledge of them.

That Craig does not pay attention to this kind of tradition is perhaps due to an inordinate focus on epistemology in his discussion of the similarity thesis. Had he looked at social and political theories, the contrast with the Romantics would not have been as striking as he believes. More generally, I think Craig has mistaken one particular version of a theory about our relation to God as the core or constitutive version. It had, I believe, a good deal more serious competition than his argument would indicate.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

It is unlikely that anyone will agree with everything in a book of this scope, and its strength lies as much in the questions and the perspectives it opens up as in its attempts to resolve these questions. The last two chapters, which I have not had the opportunity to discuss here, are especially illuminating, and offer a profound discussion of the relation between knowledge and belief. This is a fascinating book, comparable in its breadth and insight to Richard Rorty's *Mirror of Nature,* and it deserves as wide an audience.

STEPHEN GAUKROGER

University of Sydney

The Concept of Physical Law. NORMAN SWARTZ. New York: Cambridge University Press, 1985. 220 + xi p. Cloth $29.95.

This volume is devoted to a systematic defense of the "regularity" view of physical (i.e., natural) laws against the claims of those who would distinguish in some "ontological" way between generalizations which are merely "accidentally" true and those having genuine "nomic necessity." All contingent generalizations can, at best, truly state the uniformities present in the actual world. Attributions of some kind of modal force to some, alleging that they hold true not simply in the actual world but in all "physically possible" worlds as well, rest on faulty intuitions which cannot be backed up by any adequate theory. Much of the argument is framed in terms of the debate over existential assertions, rather than that over generalizations. Here the proponent of a distinction between the genuinely lawlike and the merely accidental claims a distinction between existentials that are "doomed" to falsity (since they conflict with genuine laws) and those merely false, and, hence, merely contradicting true accidental generalizations. Norman Swartz repeatedly insists that, insofar as the issue of possibility arises, it arises for all the existential assertions equally. They are, if false, "impossible" relative to some general assertion's being true. But only that kind of relative possibility is intelligible. Some further distinction in the (physical) modality of the generalizations whose truth guarantees their falsehood is merely an illusion.

What is wrong with the notion of nomic necessity? It is, Swartz says, impossible to explain what it means without invoking notions, such as "physically accessible possible worlds," that can themselves only be understood by utilizing the dubious notion of nomic neces-

0022-362X/90/8708/432–5
© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

sity with which we started. The proponent of nomic necessity fails, Swartz claims, to give us an adequate epistemological account of how we could gain epistemic access to what goes on in the alleged other physically possible worlds and what does not happen in them. More specifically interesting in the attack on this particular modality is the absence, according to Swartz, of an adequate criterion to distinguish the genuinely lawlike from the merely accidental. Logical (and, perhaps, mathematical?) assertions are characterized by a distinctive formal structure that marks them off from all other assertions, so that we can, if we want, attribute a special modal status to them. But the genuinely lawlike generalizations are not so a priori distinguishable from the mere accidental generalizations. Nor is there some clear empirical means of determining which of the generalizations true of the world have nomic status.

In the usual Humean tradition, one would expect the author to spend a fair amount of time trying to offer us at least a "psychological" account of the source of our familiar intuitions about the distinctions between generalizations that support counterfactuals and those which do not, to note one traditional alleged distinction between the laws and the mere generalizations. We do not get too much help on this from Swartz, though. He does say that the counterfactual supporting generalizations are more "pervasive" than the others, having narrower *ceteris paribus* clauses implicitly attached to them. But by itself that is not much to go on. He says that just because a piggy bank has all copper coins in it we do not infer, "If this coin were in the piggy bank it would be copper." Our ground for denying the counterfactual force, he says, is the existence of lots of piggy banks with noncopper coins in them. But that, as Swartz acknowledges, will not, by itself do. We do infer, "If this were an electron it would have a unit charge," even though electrons are all particles and lots of particles do not have unit charges.

One might expect, at this point, the usual resort of the Humean to factors not directly related to the truth conditions of the assertion to make the distinction for us. Indeed, Swartz has some extended discussion of Nicholas Rescher's[1] particular version of that sort of thesis where he advocates a version of the "role in the scientific hierarchy" model of "imputed lawlikeness." Swartz devotes his attention, though, to arguing that such a version of the grounds for the distinction between the lawlike and the accidental would not be what the proponent of nomic necessity as an "ontological" feature of the world intended. No, it would not. But, of course, the Humean line

[1] *Scientific Explanation* (New York: Free Press, 1970).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

has always been that a reconstruction of the distinction based on epistemic, pragmatic, or "psychological" features is part and parcel of an adequate account that will do justice to the real grounds of the intuitions on which the "metaphysician" relied in trying to convince us of the existence "in the world" of the structure we are now, in a sense, denying. Swartz's line of boldly saying of all false existentials that they are equally impossible of truth (since all equally incompatible with some true general assertion or other, and all such generalizations are, modally, on a par) is, perhaps, less satisfying than the Humean program that "reconstructs" the very notion of possibility, explaining (or, if you want, explaining away) our systematic intuitions instead of just denying them.

The book has three sections, "Theory," "Applications," and "The Theory Extended." The issue of free will is taken up in the second part. Here Swartz begins with the plausible regularist line that free will is compatible with lawfulness as the regularist construes it. Locutions telling us that the law "makes" the choice go as it does are misleading. It is our choosing as we do, rather, that determines what laws hold, the laws being nothing but general summaries of the particular real happenings in the world. But he then goes on and argues for unpredictability as the *sina qua non* of a free choice. Even on a surface level this is a problematic thesis, of course. The old examples of the predictable free choice of the rational agent versus the unpredictable random activity of the unfree madman come to mind.

There are problems at deeper levels, too. Swartz uses a coarse-grained notion of event (or action) so that an action could be redescribed and placed under quite different laws than the kind of generalizations connecting unique antecedent to unique action which, he says, could not have been inferred from past experience. Swartz does take this up, suggesting that, even at the lower levels, the antecedent adequate for the action will be so complex that the proper regularity could not be inferred. But, of course, the usual line would be that the antecedent is just a congeries of simple facts connected to the action, also now a collection of simple facts, by the standardly inferred physical regularities of the world. Is it just then the complexity of initial conditions (rather than the noninferribility of the appropriate regularity, as Swartz seems to suggest) that makes all actions appropriately called "free" genuinely unpredictable? Swartz also thinks that another "positive" condition on action's being free (over and above the compatibilists usual "negative" conditions, like the action being noncoerced, not the result of drugs or electrodes in the brain, etc.) along with unpredictability is that the actor not be "uniform" in his behavior. This is a tricky notion (being relative to the description

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of the action chosen). Nor is it so clear that even action under a guise that takes it as uniform should be, a fortiori, classified automatically as unfree.

At one point (144) Swartz seems to suggest that, if all events were rationally forecastable, this would provide support for the "autonomy theory," i.e., for the idea of a distinct kind of "ontological" nomic necessity. Since, he thinks, events are not universally so rationally predictable, no such support exists. But why would a world that satisfied the "Laplacean paradigm" support the "autonomy theory?" Just what is supposed to be incompatible between predictability (in principle or otherwise) of all events from inductively inferrible generalizations and knowable singular initial conditions, and the view that the regularities in question are, like all regularities, nothing but true generalizations holding in the actual world? There seems to be some implicit assumption here that, if all the regularities were so inferrible from less than an inspection of all their instances, this would indicate the truth of the necessitarian's claim that the laws "made" events go the way they go and did not just summarize how, in fact, the events do go. But why should Swartz's regularism be so threatened by such a possibility of inductive inferribility of laws and predictability of singular events?

One place where Swartz's regularism may be going too far is where (137) he suggests that, even for logical necessity, talking of the events as "complying" with the law is misleading. Instead, he says, with the use of possible-world semantics we can see that even logical truths just "take their truth from the way the world is" (138). But summarizing what goes on in "all possible worlds" is not like summarizing what goes on in the actual world. What the possible worlds *are* is constrained by the class of logical truths, unless, I suppose you really hold to a "realistic" possible-worlds way of looking at things where the actual world is just our neighborhood of reality. At this point, I think Swartz's extension of his idea that the singular facts are primary and all generalization supervenient on them gets misleading.

Many other topics are taken up: miracles and marvels from the regularity perspective, the place of statistical generalization in the regularist view, dispositions as a putative ground for natural necessity, and others. Overall the book is very well structured. The positions taken and the arguments for them are laid out with clarity. Many very trenchant replies to misguided "refutations" of regularist views are forcefully and convincingly presented. The book is a welcome contribution to the recurrent debate on the nature of lawlike generality.

University of Michigan

LAWRENCE SKLAR

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NOTES AND NEWS

A conference on *Rationality, Morality, and Self-interest*, dedicated to the memory of Mark Overvold (who died in 1988), will be held from October 18–20 at St. Olaf College. Information on the program and accommodations may be obtained by writing Frederick Stoutland, Phil. Dept., St. Olaf College, Northfield, MN 55057.

The Centre for Philosophy and Public Affairs at the University of St. Andrews announces its Research Fellowships for the academic session 1991–92. The fellowships are intended primarily, though not exclusively, for philosophers and political theorists on study leave from their own universities or colleges. The deadline for application is November 16. Further details are available from the Director, Dr. John Haldane, Centre for Philosophy and Public Affairs, University of St. Andrews, North Street, St. Andrews, Fife—KY16 9AJ, Scotland, U.K.

Harvard University, The Program in Ethics and the Professions, invites applications for Fellowships in Professional Ethics. Six Fellowships will be awarded in 1991–92 to outstanding teachers and scholars who wish to develop their competence to address ethical issues in schools of business, government, law, and medicine. Fellows will participate in the seminar of the Program, attend courses in one of the professional schools or in the Graduate School of Arts and Sciences, and conduct their own research on professional ethics. Application deadline is January 7, 1991. For further information, you may write to Jean McVeigh, Administrator, The Program in Ethics and the Professions, Kennedy School of Government, Harvard University, 79 Kennedy Street, Cambridge, MA 02138. Harvard University also announces its Andrew W. Mellon Faculty Fellowships in the Humanities, for nontenured, experienced junior scholars who have completed, at the time of appointment, at least two years postdoctoral teaching as college or university faculty in the humanities—usually as assistant professors. Applications due November 1. For particulars and application procedures you may write to: Dr. Richard M. Hunt, Harvard Univ. Mellon Faculty Fellowships, Lamont Library 202, Cambridge, MA 02138.

The Philosophy Department at the University of Rochester is happy to announce that the prize winner in the first year of the Colin and Alisa Turbayne International Berkeley Essay Prize Competition is Philip Cummins of the University of Iowa for his essay, "Berkeley's Manifest Qualities Thesis." The next deadline for submitting papers is November 1. Essays should be new and unpublished and should be written in English, not to exceed 5000 words in length. Submissions may be sent to Ralf Meerbote, Phil. Dept., Univ. of Rochester, Rochester, NY 14627.

The Department of Philosophy at California State University/Northridge is pleased to announce the appointment of James E. Maffie as Assistant Professor.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The Department of Philosophy at Clark University is pleased to announce the appointment of Judith Wagner DeCew as Associate Professor.

The Department of Philosophy of Indiana University/Bloomington is pleased to announce the appointments of Leah Savion and Zeno Swijtink as Assistant Professors. Additionally, Dennis Senchuk will become a full-time member of the Department of Philosophy.

The Philosophy Department of the University of Massachusetts/Amherst is pleased to announce the appointments of Phillip Bricker as Associate Professor and John G. Robison as Head of the Department.

The Department of Philosophy at the University of California/Riverside is pleased to announce the appointments of Michael Tye as Professor, and Jonathan Wilwerding and Mark Ravizza as Assistant Professors.

The Department of Philosophy at the University of Colorado/Boulder is pleased to announce the appointment of Luc Bovens as Assistant Professor.

The Department of Philosophy at the University of Tennessee/Knoxville is pleased to announce the appointment of Lisa Parker as Assistant Professor.

The Department of Philosophy at Virginia Commonwealth University is pleased to announce the appointment of Tony Ellis as full Professor. Tony will continue his editorship of *Philosophical Books* at Virginia Commonwealth University.

NEW BOOKS: TRANSLATIONS

ALEXANDER OF APHRODISIAS: *On Aristotle's Metaphysics 1.* W. E. DOOLEY, trans. Ithaca: Cornell, 1989. 240 p. Cloth $41.50.

ALEXY, ROBERT: *A Theory of Legal Argumentation: The Theory of Rational Discourse as Theory of Legal Justification.* RUTH ADLER & NEIL MACCORMICK, trans. New York: Oxford, 1989. xiii, 323 p. Cloth $69.00.

ARISTOTLE: *Prior Analytics.* ROBIN SMITH, trans. Indianapolis: Hackett, 1989. xxxi, 262 p. Cloth $27.50.

BITTNER, RUDIGER: *What Reason Demands.* THEODORE TALBOT, trans. New York: Cambridge, 1989. ix, 198 p. Cloth $37.50, paper $12.95.

CANGUILHEM, GEORGES: *The Normal and the Pathological.* CAROLYN R. FAWCETT, trans. New York: Zone, 1989. 327 p. Cloth $24.50.

DELEUZE, GILLES: *The Logic of Sense.* MARK LESTER, trans. New York: Columbia, 1990. xiv, 393 p. Cloth $45.00; *Cinema 2: The Time Image.* HUGH TOMLINSON & ROBERT GALETA, trans. Minneapolis: Minnesota UP, 1989. xvii, 344 p. Cloth $39.50, paper $14.95.

DERRIDA, JACQUES: *Of Spirit: Heidegger and the Question.* GEOFFREY BENNINGTON & RACHEL BOWLBY, trans. Chicago: University Press, 1989. vii, 139 p. Cloth $19.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

DESCARTES, RENE: *The Passions of the Soul*. STEPHEN VOSS, trans. Indianapolis: Hackett, 1989. xxv, 165 p. Cloth $22.50, paper $4.95.

FARIAS, VICTOR: *Heidegger and Nazism*. GABRIEL R. RICCI & PAUL BURREL, trans. Philadelphia: Temple, 1989. xxi, 349 p. Cloth $29.95.

HABERMAS, JÜRGEN: *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. THOMAS BURGER, trans. Cambridge: MIT, 1989. xix, 301 p. Cloth $35.00.

HUSSERL, EDMUND: *Aufsatze und Vortrage (1922–1937)*. Boston: Kluwer, 1989. xxx, 332 p. Cloth $142.00; *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy, Second Book: Studies in the Phenomenology of Constitution*. RICHARD ROJCEWICZ and ANDRE SCHUWER, trans. Boston: Kluwer, 1989. xix, 439 p. Cloth $134.00.

JAUSS, HANS ROBERT: *Question and Answer: Forms of Dialogic Understanding*. MICHAEL HAYS, trans. Minneapolis: Minnesota UP, 1989. x, 283 p. Cloth $45.00, paper $16.95.

KOLAKOWSKI, LESZEK: *The Presence of Myth*. ADAM CZERNIAWSKI, trans. Chicago: University Press, 1989. xii, 138 p. Cloth $19.95.

LACOUE-LABARTHE, PHILIPPE: *Typography: Mimesis, Philosophy, Politics*. CHRISTOPHER FYNSK, et al. eds. & trans. Cambridge: Harvard, 1989. x, 308 p. Cloth $35.00.

LUHMANN, NIKLAS: *Ecological Communication*. JOHN BEDNARZ, Jr., Trans. Chicago: University Press, 1989. xviii, 187 p. Cloth $34.95.

MONTESQUIEU: *The Spirit of the Laws*. ANNE COHLER, BASIA MILLER, & HAROLD STONE, eds. & trans. New York: Cambridge, 1989. xlvii, 757 p. Cloth $59.50, paper $16.95.

MORE: *Utopia*. GEORGE M. LOGAN & ROBERT M. ADAMS, eds. & trans. New York: Cambridge, 1989. xxxv, 137 p. Cloth $29.95, paper $6.95.

NAESS, ARNE: *Ecology, Community and Lifestyle: Outline of an Ecosophy*. DAVID ROTHENBERG, trans. New York: Cambridge, 1989. xii, 223 p. Cloth $44.50.

NIETZSCHE, FRIEDRICH: *Unmodern Observations*. WILLIAM ARROWSMITH, ed. & trans. New Haven: Yale, 1990. xix, 402 p. Cloth $40.00.

PLATO: *Symposium*. TOM GRIFFITH, trans. Berkeley: California UP, 1989. 144 p. Cloth $25.00.

PLATO: *Symposium*. ALEXANDER NEHAMAS & PAUL WOODRUFF, trans. Indianapolis: Hackett, 1989. xxvii, 80 p. Cloth $17.50, paper $3.45.

SCHOPENHAUER, ARTHUR: *Manuscript Remains in Four Volumes*. ARTHUR HUBSCHER, ed., E. F. J. PAYNE, trans. New York: Berg, 1989. xx, 760 p. Cloth $85.00.

SAFRANSKI, RUDIGER: *Schopenhauer and the Wild Years of Philosophy*. EWALD OSERS, trans. Cambridge: Harvard, 1990. vii, 385 p. Cloth $30.00.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVII, NO. 9, SEPTEMBER 1990

NATIONAL SELF-DETERMINATION*

IN the controversy-ridden fields of international law and international relations, the widespread recognition of the existence of national rights to self-determination provides a welcome point of agreement. Needless to say, the core consensus is but the eye of a raging storm concerning the precise definition of the right, its content, its bearers, and the proper means for its implementation. This paper will not address such questions, though indirectly it may help with their investigation. Its concern is with the moral justification of the case for national self-determination. Its purpose is critical and evaluative, its subject lies within the morality of international relations rather than within international law and international relations proper.

It is assumed throughout that states and international law should recognize such a right only if there is a sound moral case for it. This does not mean that international law should mirror morality. Its concern is with setting standards that enjoy the sort of clarity required to make them the foundations of international relations between states and fit for recognition and enforcement through international organs. These concerns give rise to special considerations that should be fully recognized in the subtle process of applying moral principles to the law. The derivation of legal principles from moral premises is never a matter of copying morality into law. Still, the justification of the law rests ultimately on moral considerations, and therefore those considerations should also help shape the contours of legal principles. That is why the conclusions of this paper bear on controversies concerning the proper way in which the law on this subject should develop, even though such issues are not here discussed directly.

*We are grateful to Lea Brilmayer, Moshe Halbertal, David Heyd, and the editors of the Journal for helpful comments on an earlier draft.

0022-362X/90/8709/439–461

© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Moral inquiry is sometimes understood in a utopian manner, i.e., as an inquiry into the principles that should prevail in an ideal world. It is doubtful whether this is a meaningful enterprise, but it is certainly not the one we are engaged in here. We assume that things are roughly as they are, especially that our world is a world of states and of a variety of ethnic, national, tribal, and other groups.[1] We do not question the justification for this state of affairs. Rather, we ask whether, given that this is how things are and for as long as they remain the same, a moral case can be made in support of national self-determination.

I. ISOLATING THE ISSUE

The core content of the claim to be examined is that there is a right to determine whether a certain territory shall become, or remain, a separate state (and possibly also whether it should enjoy autonomy within a larger state). The idea of national self-determination or (as we shall refer to it in order to avoid confusion) the idea of self-government encompasses much more. The value of national self-government is the value of entrusting the general political power over a group and its members to the group. If self-government is valuable then it is valuable that whatever is a proper matter for political decision should be subject to the political decision of the group in all matters concerning the group and its members. The idea of national self-government, in other words, speaks of groups determining the character of their social and economic environment, their fortunes, the course of their development, and the fortunes of their members by their own actions, i.e., by the action of those groups, in as much as these are matters which are properly within the realm of political action.[2] Given the current international state system, in which politi-

[1] This fact is doubly relevant. It is a natural fact about our world that it is a populated world with no unappropriated lands. It is a social and a moral fact that it is a world of nations, tribes, peoples, etc., that is, that people's perception of themselves and of others and their judgments of the opportunities and the responsibilities of life are shaped, to an extent, by the existence of such groups and their membership of them. It may be meaningful to claim that our views regarding national self-determination apply only to a populated world like ours. One may point to different principles that would prevail in a world with vast unoccupied fertile lands. Such speculation is utopian but it may serve to highlight some of the reasons for the principles that apply in our condition. To speculate concerning a reality different from ours in its basic social and moral constitution is pointless in a deeper way. Such social facts are constitutive of morality. Their absence undercuts morality's very foundations. We could say that under such changed conditions people will have normative beliefs and will be guided by some values. But they are not ones for which we can claim any validity.

[2] This qualification is to take account of the fact that, according to doctrines of limited government, certain matters are outside the realm of politics, and no political action regarding them may be undertaken.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

cal power rests, in the main, with sovereign states,[3] the right to determine whether a territory should be an independent state is quite naturally regarded as the main instrument for realizing the ideal of self-determination. Consideration of this right usually dominates all discussions of national self-determination. To examine the justification of the right is the ultimate purpose of this article. But we shall continuously draw attention to the fact that, as we shall try to show, the right of self-determination so understood is not ultimate, but is grounded in the wider value of national self-government, which is itself to be only instrumentally justified.

The next section deals with the nature of the groups that might be the subject of such a right. Section III considers what value, if any, is served by the enjoyment of political independence by such groups. Section IV examines the case for conceding that there is a moral right to self-determination. This examination may lead to revising our understanding of the content of the right. It may reveal that moral considerations justify only a narrower right, or that the argument that justifies the right warrants giving it a wider scope. But the core as identified here will provide the working base from which to launch the inquiry.

Before we start, a few words about this way of identifying the problem may be in place. In two ways the chosen focus of our examination is narrower than many discussions of self-determination in international relations. First, we disregard the claims made, typically by third-world countries, in the name of self-determination, against the economic domination of multinational companies, the World Bank, or against powerful regional or world powers. The considerations canvassed in this paper are relevant to such issues, but fall short of directly tackling them. To be complete, a discussion of a right must examine both its grounds and its consequences. This paper is concerned mostly with the grounds for the right of self-determination. It asks the question: Who has the right and under what conditions is it to be exercised? It does not go into the question of the consequences of the right beyond the assumption, already stated, that it is a right that a territory be a self-governing state. A good deal of the current turmoil in international law, and international relations, has to do with the exploration of that last notion. What is entailed by the fact that a state is a sovereign, self-governing, entity?

[3] Among the exceptions to this rule are the slowly growing importance of supernational, especially regional, associations, such as the European Community, the growth of a doctrine of sovereignty limited by respect for fundamental human rights, and the continuing (usually thinly veiled) claims of some states that they are not bound by the international law regarding the sovereignty of states.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The claims that economic domination violate the right to self-determination belong to that discussion. The conclusions of this paper provide part of the grounds by which such claims are to be settled. But we do not propose to pursue this question here.

Second, claims of self-determination are invariably raised whenever one state invades and occupies another, or a territory belonging to another. Yet it is important to distinguish between the wrongness of military invasion or occupation, and the rights available against it, and the right (whatever it may turn one to be) to self-determination. In a word, the latter is a source of title, whereas the former is a possessory right based largely on public-order considerations. Any legal system, international law not excluded, recognizes certain ways as legitimate ways of solving disputes, and outlaws others. Subject to the exceptions of legitimate self-defense and self-help, the use of violence is forbidden. Violation of that prohibition gives rise to a right to have the *status quo ante* restored, before the deeper sources of the dispute between the parties are examined; that is, regardless of the soundness of one's title to a territory, one may not use force to occupy it. This is why the right to recover a territory lost by force is a possessory right. It does not depend on the ultimate soundness of one's title, and that is why it was said to be based on public-order considerations. A large part of its justification is in the need to establish that the proper means of dispute resolution be the only ones resorted to.

Not surprisingly, invocation of this possessory right is, however, accompanied by a claim of good title (the merits of which are not immediately relevant). The underlying title is often the right to self-determination. Hence the temptation to confuse the two. But notice that, apart from the different justificatory foundations, the two are far from identical in consequence. They merely overlap. The claims of a people who have been for many years ruled by another cannot be based on the possessory right that applies only against a recent occupier. On the other hand, the occupation of portions of Antarctica, or of some uninhabited island, do violate the possessory right, but not the right of self-determination. The latter is that of the inhabitants, and does not apply when there are no inhabitants.[4]

II. GROUPS

Assuming that self-determination is enjoyed by groups, what groups qualify? Given that the right is normally attributed to peoples or

[4] The substantive right protected indirectly by the possessory right in cases of this kind is one of the other rights providing a title in a territory. The right to self-determination is only one of the possible sources of title.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

nations, it is tempting to give that as the answer and concentrate on characterizing "peoples" or "nations." The drawbacks of this approach are two: it assumes too much and it poses problems that may not require a solution.

It is far from clear that peoples or nations rather than tribes, ethnic groups, linguistic, religious, or geographical groups are the relevant reference group. What is it that makes peoples particularly suited to self-determination? The right concerns determination whether a certain territory shall be self-governing or not. It appears to affect most directly the residents of a territory, and their neighbors. If anyone, then residents of geographical regions seem intuitively to be the proper bearers of the right. Saying this does not get us very far. It does not help in identifying the residents of which regions should qualify. To be sure, this is the crucial question. But even posing it in this way shows that the answer, "the largest regions inhabited by one people or nation," is far from being the obvious answer.

We have some understanding of the benefits self-government might bring. We need to rely on this in looking for the characteristics that make groups suitable recipients of those benefits. We want, in other words, to identify groups by those characteristics which are relevant to the justification of the right. If it turns out that those do not apply to peoples or nations, we shall have shown that the right to self-determination is misconceived and, as recognized in international law, unjustified. Alternatively, the groups identified may encompass peoples (or some peoples) as well as other groups. This will provide a powerful case for redrawing the boundaries of the right. Either way we shall be saved much argument concerning the characterization of nations which, interesting as it is in itself, is irrelevant to our purpose.

Having said that, it may be useful to take nations and peoples as the obvious candidates for the right. We need not worry about their defining characteristics. But we may gain insight by comparing them with groups, e.g., the fiction-reading public, or Tottenham Football Club supporters, which obviously do not enjoy such a right. Reflection on such examples suggests six characteristics that in combination are relevant to a case for self-determination.

1. The group has a common character and a common culture that encompass many, varied and important aspects of life, a culture that defines or marks a variety of forms or styles of life, types of activities, occupations, pursuits, and relationships. With national groups we expect to find national cuisines, distinctive architectural styles, a common language, distinctive literary and artistic traditions, national

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

music, customs, dress, ceremonies and holidays, etc. None of these is necessary. They are but typical examples of the features that characterize peoples and other groups that are serious candidates for the right to self-determination. They have pervasive cultures, and their identity is determined at least in part by their culture. They possess cultural traditions that penetrate beyond a single or a few areas of human life, and display themselves in a whole range of areas, including many which are of great importance for the well-being of individuals.

2. The correlative of the first feature is that people growing up among members of the group will acquire the group culture, will be marked by its character. Their tastes and their options will be affected by that culture to a significant degree. The types of careers open to one, the leisure activities one learned to appreciate and is therefore able to choose from, the customs and habits that define and color relations with strangers and with friends, patterns of expectations and attitudes between spouses and among other members of the family, features of lifestyles with which one is capable of empathizing and for which one may therefore develop a taste—all these will be marked by the group culture.

They need not be indelibly marked. People may migrate to other environments, shed their previous culture, and acquire a new one. It is a painful and slow process, success in which is rarely complete. But it is possible, just as it is possible that socialization will fail and one will fail to be marked by the culture of one's environment, except negatively, to reject it. The point made is merely the modest one that, given the pervasive nature of the culture of the groups we are seeking to identify, their influence on individuals who grow up in their midst is profound and far-reaching. The point needs to be made in order to connect concern with the prosperity of the group with concern for the well-being of individuals. This tie between the individual and the collective is at the heart of the case for self-determination.

As one would expect, the tie does not necessarily extend to all members of the group, and failure of socialization is not the only reason. The group culture affects those who grow up among its members, be they members or not. But to say this is no more than to point to various anomalies and dilemmas that may arise. Most people live in groups of these kinds, so that those who belong to none are denied full access to the opportunities that are shaped in part by the group's culture. They are made to feel estranged and their chances to have a rewarding life are seriously damaged. The same is true of people who grow up among members of a group so that they absorb

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

its culture, but are then denied access to it because they are denied full membership of the group.

Nothing in the above presupposes that groups of the kind we are exploring are geographically concentrated, let alone that their members are the only inhabitants of any region. Rather, by drawing on the transmission of the group culture through the socialization of the young, these comments emphasize the historical nature of the groups with which we are concerned. Given that they are identified by a common culture, at least in part, they also share a history, for it is through a shared history that cultures develop and are transmitted.

3. Membership in the group is, in part, a matter of mutual recognition. Typically, one belongs to such groups if, among other conditions, one is recognized by other members of the group as belonging to it. The other conditions (which may be the accident of birth or the sharing of the group culture, etc.) are normally the grounds cited as reasons for such recognition. But those who meet those other conditions and are yet rejected by the group are at best marginal or problematic members of it. The groups concerned are not formal institutionalized groups, with formal procedures of admission. Membership in them is a matter of informal acknowledgment of belonging by others generally, and by other members specifically. The fiction-reading public fails our previous tests. It is not identified by its sharing a wide-ranging pervasive culture. It also fails the third test. To belong to the fiction-reading public all we have to do is to read fiction. It does not matter whether others recognize us as fiction-reading.[5]

4. The third feature prepares the way for, and usually goes hand in hand with, the importance of membership for one's self-identification. Consider the fiction-reading public again. It is a historically significant group. Historians may study the evolution of the fiction-reading public, how it spread from women to men, from one class to others, from reading aloud in small groups to silent reading, from reliance on libraries to book buying, etc.; how it is regarded as important to one's qualification as a cultured person in one country, but not in another; how it furnishes a common topic of conversation in some classes but not in others; how belonging to the group is a mark of political awareness in some countries, while being a sign of escapist retreat from social concerns in another.

Such studies will show, however, that it is only in some societies

[5] The fiction-reading public can take the character of a literary elite with mutual recognition as part of its identity. The importance of "acceptability" in such groups has often been noted and analyzed.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that the existence of these features of the fiction-reading public is widely known. For the most part, one can belong to the group without being aware that one is a typical reader, that one's profile is that of most readers. Sometimes this is a result of a mistaken group image's being current in that society. Our concern is rather with those cases where the society lacks any very distinct image of that group. This indicates that, in such societies, membership of that group does not have a highly visible social profile. It is not one of the facts by which people pigeonhole each other. One need not be aware who, among people one knows, friends, acquaintances, shopkeepers one patronizes, one's doctor, etc., shares the habit. In such societies, membership of the fiction-reading public is not highly visible, that is, it is not one of the things one will normally know about people one has contact with, one of the things that identify "who they are." But it happens in some countries that membership of the reading public becomes a highly visible mark of belonging to a social group, to the intelligentsia, etc. In such countries, talk of the recently published novel becomes a means of mutual recognition.

One of the most significant facts differentiating various football cultures is whether they are cultures of self-recognition: whether identification as a fan or supporter of this club or that is one of the features that are among the main markers of people in the society. The same is true of occupational groups. In some countries, membership is highly visible and is among the primary means of pigeonholing people, of establishing "who they are"; in others, it is not.

Our concern is with groups, membership of which has a high social profile, that is, groups, membership of which is one of the primary facts by which people are identified, and which form expectations as to what they are like, groups membership of which is one of the primary clues for people generally in interpreting the conduct of others. Since our perceptions of ourselves are in large measure determined by how we expect others to perceive us, it follows that membership of such groups is an important identifying feature for each about himself. These are groups, members of which are aware of their membership and typically regard it as an important clue in understanding who they are, in interpreting their actions and reactions, in understanding their tastes and their manner.

5. Membership is a matter of belonging, not of achievement. One does not have to prove oneself, or to excel in anything, in order to belong and to be accepted as a full member. To the extent that membership normally involves recognition by others as a member, that recognition is not conditional on meeting qualifications that indicate any accomplishment. To be a good Irishman, it is true, is an

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

achievement. But to be an Irishman is not. Qualification for membership is usually determined by nonvoluntary criteria. One cannot choose to belong. One belongs because of who one is. One can come to belong to such groups, but only by changing, e.g., by adopting their culture, changing one's tastes and habits accordingly—a very slow process indeed. The fact that these are groups, membership of which is a matter of belonging and not of accomplishment, makes them suitable for their role as primary foci of identification. Identification is more secure, less liable to be threatened, if it does not depend on accomplishment. Although accomplishments play their role in people's sense of their own identity, it would seem that at the most fundamental level our sense of our own identity depends on criteria of belonging rather than on those of accomplishment. Secure identification at that level is particularly important to one' well-being.

6. The groups concerned are not small face-to-face groups, members of which are generally known to all other members. They are anonymous groups where mutual recognition is secured by the possession of general characteristics. The exclusion of small groups from consideration is not merely ad hoc. Small groups that are based on personal familiarity of all with all are markedly different in the character of their relationships and interactions from anonymous groups. For example, given the importance of mutual recognition to members of these groups, they tend to develop conventional means of identification, such as the use of symbolic objects, participation in group ceremonies, special group manners, or special vocabulary, which help quickly to identify who is "one of us" and who is not.

The various features we listed do not entail each other but they tend to go together. It is not surprising that groups with pervasive cultures will be important in determining the main options and opportunities of their members, or that they will become focal points of identification, etc. The way things are in our world, just about everyone belongs to such a group, and not necessarily to one only. Membership is not exclusive and many people belong to several groups that answer to our description. Some of them are rather like national groups, e.g., tribes or ethnic groups. Others are very different. Some religious groups meet our conditions, as do social classes, and some racial groups. Not all religions or racial groups did develop rich and pervasive cultures. But some did and those qualify.

III. THE VALUE OF SELF-GOVERNMENT

(A) *The Value of Encompassing Groups.* The description of the relevant groups in the preceding section may well disappoint the reader. Some will be disappointed by the imprecise nature of the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

criteria provided. This would be unjustified. The criteria are not meant to provide operational legal definitions. As such they clearly would not do. Their purpose is to pick on the features of groups which may explain the value of self-determination. As already mentioned, the key to the explanation is in the importance of these groups to the well-being of their members. This thought guided the selection of the features. They are meant to assist in identifying that link. It is not really surprising that they are all vague matters of degree, admitting of many variants and many nuances. One is tempted to say "that's life." It does not come in neatly parceled parts. While striving to identify the features that matter, we have to recognize that they come in many shapes, in many shades, and in many degrees rife with impurities in their concrete mixing.

A more justified source of disappointment is the suspicion that we have cast the net too wide. Social classes clearly do not have a right to self-determination. If they meet the above conditions then those conditions are at best incomplete. Here we can only crave the reader's patience. We tried to identify the features of groups which help explain the value of self-determination. These may apply not only beyond the sphere in which the right is commonly recognized. They may apply to groups that really should not possess it for other reasons yet to be explored.

The defining properties of the groups we identified are of two kinds. On the one hand, they pick out groups with pervasive cultures; on the other, they focus on groups, membership of which is important to one's self-identity. This combination makes such groups suitable candidates for self-rule. Let us call groups manifesting the six features *encompassing groups*. Individuals find in them a culture which shapes to a large degree their tastes and opportunities, and which provides an anchor for their self-identification and the safety of effortless secure belonging.

Individual well-being depends on the successful pursuit of worthwhile goals and relationships. Goals and relationships are culturally determined. Being social animals means not merely that the means for the satisfaction of people's goals are more readily available within society. More crucially it means that those goals themselves are (when one reaches beyond what is strictly necessary for biological survival) the creatures of society, the products of culture. Family relations, all other social relations between people, careers, leisure activities, the arts, sciences, and other obvious products of "high culture" are the fruits of society. They all depend for their existence on the sharing of patterns of expectations, on traditions preserving implicit knowledge of how to do what, of tacit conventions regarding what is part of this

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

or that enterprise and what is not, what is appropriate and what is not, what is valuable and what is not. Familiarity with a culture determines the boundaries of the imaginable. Sharing in a culture, being part of it, determines the limits of the feasible.

It may be no more than a brute fact that our world is organized in a large measure around groups with pervasive cultures. But it is a fact with far-reaching consequences. It means, in the first place, that membership of such groups is of great importance to individual well-being, for it greatly affects one's opportunities, one's ability to engage in the relationships and pursuits marked by the culture. Secondly, it means that the prosperity of the culture is important to the well-being of its members. If the culture is decaying, or if it is persecuted or discriminated against, the options and opportunities open to its members will shrink, become less attractive, and their pursuit less likely to be successful.

It may be no more than a brute fact that people's sense of their own identity is bound up with their sense of belonging to encompassing groups and that their self-respect is affected by the esteem in which these groups are held. But these facts, too, have important consequences. They mean that individual dignity and self-respect require that the groups, membership of which contributes to one's sense of identity, be generally respected and not be made a subject of ridicule, hatred, discrimination, or persecution.

All this is mere common sense, and is meant to be hedged and qualified in the way our common understanding of these matters is. Of course, strangers can participate in activities marked by a culture. They are handicapped, but not always very seriously. Of course, there are other determinants of one's opportunities, and of one's sense of self-respect. Membership of an encompassing group is but one factor. Finally, one should mention that groups and their culture may be pernicious, based on exploitation of people, be they their members or not, or on the denigration and persecution of other groups. If so, then the case for their protection and flourishing is weakened, and may disappear altogether.

Having regard for this reservation, the case for holding the prosperity of encompassing groups as vital for the prosperity of their members is a powerful one. Group interests cannot be reduced to individual interests. It makes sense to talk of a group's prospering or declining, of actions and policies as serving the group's interest or of harming it, without having to cash this in terms of individual interests. The group may flourish if its culture prospers, but this need not mean that the lot of its members or of anyone else has improved. It is in the interest of the group to be held in high regard by others, but it

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

does not follow that, if an American moon landing increases the world's admiration for the United States, Americans necessarily benefit from this. Group interests are conceptually connected to the interests of their members but such connections are nonreductive and generally indirect. For example, it is possible that what enhances the interest of the group provides opportunities for improvement for its members, or that it increases the chance that they will benefit.

This relative independence of group interest is compatible with the view that informs this article: that the moral importance of the group's interest depends on its value to individuals. A large decline in the fortunes of the group may, e.g., be of little consequence to its members. There is no a priori way of correlating group interest with that of its members or of other individuals. It depends on the circumstances of different groups at different times. One clear consequence of the fact that the moral significance of a group's interest is in its service to individuals is the fact that it will depend, in part, on the size of the group. The fortunes of a larger group may be material to the well-being of a larger number of people. Other things being equal, numbers matter.

(B) *The Instrumental Case.* Does the interest of members in the prosperity of the group establish a right to self-determination? Certainly not, at least not yet, not without further argument. For one thing we have yet to see any connection between the prosperity of encompassing groups and their political independence. The easiest connection to establish under certain conditions is an instrumental one. Sometimes the prosperity of the group and its self-respect are aided by, sometimes they may be impossible to secure without, the group's enjoying political sovereignty over its own affairs. Sovereignty enables the group to conduct its own affairs in a way conducive to its prosperity.[6] There is no need to elaborate the point. It depends on historical conditions. Hence the prominence of a history of persecution in most debates concerning self-determination. But a history of persecution is neither a necessary nor a sufficient condition for the instrumental case for self-government. It is not a necessary condition, because persecution is not the only reason why the groups may suffer without independence. Suffering can be the result of neglect or ignorance of or indifference to the prosperity of a minority group by the majority. Such attitudes may be so well entrenched that there is no realistic prospect of changing them.

[6] This is not meant to suggest that there are not often drawbacks to self-rule. They will be considered below.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Persecution is not a sufficient condition, for there may be other ways to fight and overcome persecution and because whatever the advantages of independence it may, in the circumstances, lead to economic decline, cultural decay, or social disorder, which only make their members worse off. Besides, as mentioned above, pernicious groups may not deserve protection, especially if it will help them to pursue repressive practices with impunity. Finally, there are the interests of nonmembers to be considered. In short, the instrumental argument (as well as others) for self-government is sensitive to counterarguments pointing to its drawbacks, its cost in terms of human well-being, possible violations of human rights, etc.

We shall return to these issues below. First, let us consider the claim that the instrumental argument trivializes the case for self-government by overlooking its intrinsic value. Of the various arguments for the intrinsic value of self-government which have been and can be advanced, we examine one which seems the most promising.

(C) *An Argument for the Intrinsic Value of Self-government.* The argument is based on an extension of individual autonomy or of self-expression (if that is regarded as independently valuable). The argument unravels in stages: (1) people's membership of encompassing groups is an important aspect of their personality, and their well-being depends on giving it full expression; (2) expression of membership essentially includes manifestation of membership in the open, public life of the community; (3) this requires expressing one's membership in political activities within the community. The political is an essential arena of community life, and consequently of individual well-being; (4) therefore, self-government is inherently valuable, it is required to provide the group with a political dimension.

The first premise is unexceptionable. So is the second, though an ambiguity might be detected in the way it is often understood. Two elements need separating. First, given the importance of membership to one's well-being, it is vital that the dignity of the group be preserved. This depends, in part, on public manifestations of respect for the group and its culture, and on the absence of ridicule of the group, etc., from the public life of the society of which one is a member. One should not have to identify with or feel loyalty to a group that denigrates an encompassing group to which one belongs. Indeed, one should not have to live in an environment in which such attitudes are part of the common culture. Second, an aspect of well-being is an ability to express publicly one's identification with the group and to participate openly in its public culture. An encompassing group is centered on mutual recognition and is inevitably a

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

 Digitized by Arya Samaj Foundation Chennai and eGangotri

group with a public culture. One cannot enjoy the benefits of membership without participation in its public culture, without public participation in its culture.

Both elements are of great importance. Both indicate the vital role played by public manifestations of group culture and group membership among the conditions of individual well-being. To the extent that a person's well-being is bound up with his identity as a member of an encompassing group it has an important public dimension. But that dimension is not necessarily political in the conventional narrow sense of the term. Even where it is, its political expression does not require a political organization whose boundaries coincide with those of the group. One may be politically active in a multinational, multicultural polity.

Here supporters of the argument for the intrinsic value of self-government may protest. The expression of membership in the political life of the community, they will say, involves more than its public expression. It involves the possibility of members of an encompassing group participating in the political life of their state, and fighting in the name of group interests in the political arena. Such actions, they will insist, may be not only instrumentally valuable to the group, but intrinsically important to its politically active members. They are valuable avenues of self-fullfilment. These points, too, have to be readily admitted. There is no reason to think that everyone must take part in politics, or else his or her development is stunted and personality or life are deficient. In normal times, politics is but an option that people may choose to take or to leave alone. Although its availability is important, for its absence deprives people of valuable opportunities, its use is strictly optional. Even if it is possible to argue that one's personal well-being requires some involvement with larger groups, and the avoidance of exclusive preoccupation with one's own affairs and those of one's close relations or friends, that involvement can take nonpolitical forms, such as activity in a social club, interest in the fortunes of the arts in one's region, etc.

Politics is no more than an option, though this is true in normal times only. In times of political crises that have moral dimensions, it may well be the duty of everyone to stand up and be counted. In Weimar, Germans had a moral duty to become politically involved to oppose Nazism. There are many other situations where an apolitical attitude is not morally acceptable. But all of them are marked by moral crises. In the absence of crisis there is nothing wrong in being nonpolitical.

Having said this, we must repeat that the option of politics must remain open, and with it the option of fighting politically for causes

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

to do with the interests of one's encompassing groups. But there is nothing here to suggest that this should be done in a political framework exclusive to one's group or dominated by it. There is nothing wrong with multinational states, in which members of the different communities compete in the political arena for public resources for their communities. Admittedly, prejudice, national fanaticism, etc., sometimes make such peaceful and equitable sharing of the political arena impossible. They may lead to friction and persecution. This may constitute a good argument for the value of self-government, but it is an instrumental argument of the kind canvassed above. There is nothing in the need for a public or even a political expression of one's membership of an encompassing group which points to an intrinsic value of self-government.

(D) *The Subjective Element.* In an indirect way, the attempt to argue for the intrinsic value of self-government does point to the danger of misinterpreting the instrumental approach to the question. First, the argument does not deny the intrinsic value of the existence of the political option as a venue for activity and self-expression to all (adult) members of society. We are not advocating a purely instrumentalist view of politics generally. The intrinsic value to individuals of the political option does not require expression in polities whose boundaries coincide with those of encompassing groups. That is the only point argued for above.

Second, the pragmatic, instrumentalist character of the approach advocated here should not be identified with an aggregating impersonal consequentialism. Some people tend to associate any instrumentalist approach with images of a bureaucracy trading off the interest of one person against that of another on the basis of some cost-benefit analysis designed to maximize overall satisfaction; a bureaucracy, moreover, in charge of determining for people what is really good for them, regardless of their own views of the matter. Nothing of the kind should be countenanced. Of course, conflicts among people's interests do arise, and call for rational resolution that is likely to involve sacrificing some interests of some people for the sake of others. Such conflicts, however, admit of a large degree of indeterminacy, and many alternative resolutions may be plausible or rational. In such contexts, talking of maximization, with its connotations of comparability of all options, is entirely out of place.

Furthermore, nothing in the instrumentalist and pragmatic nature of our approach should be allowed to disguise its sensitivity to subjective elements, its responsiveness to the perceptions and sensibilities of the people concerned. To a considerable extent, what matters

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

is how well people feel in their environment: Do they feel at home in it or are they alienated from it? Do they feel respected or humiliated? etc. This leads to a delicate balance between "objective" factors and subjective perceptions. On the one hand, when prospects for the future are concerned, subjective perceptions of danger and likely persecution, etc., are not necessarily to be trusted. These are objective issues on which the opinion of independent spectators may be more reliable than that of those directly involved. On the other hand, the factual issue facing the independent spectators is how people will respond to their conditions, what will be their perceptions, their attitudes to their environment, to their neighbors, etc. Even a group that is not persecuted may suffer many of the ills of real persecution if it feels persecuted. That its perceptions are mistaken or exaggerated is important in pointing to the possibility of a different cure: removing the mistaken perception. But that is not always possible, and up to a point in matters of respect, identification, and dignity, subjective responses, justified or not, are the ultimate reality so far as the well-being of those who have them is concerned.

IV. A RIGHT TO SELF-DETERMINATION

It may seem that the case for self-government establishes a right to self-determination. That is, it establishes the reasons for the right sort of group, an encompassing group, to determine that a territory shall be self-governing. But things are not that simple. The case for self-government shows that sometimes, under certain conditions, it is best that the political unit be roughly an encompassing group. A group's right to self-determination is its right to determine that a territory be self-governing, regardless of whether the case for self-government, based on its benefits, is established or not. In other words, the right to self-determination answers the question 'who is to decide?', not 'what is the best decision?'. In exercising the right, the group should act responsibly in light of all the considerations we mentioned so far. It should, in particular, consider not only the interests of its members but those of others who may be affected by its decision. But if it has the right to decide, its decision is binding even if it is wrong, even if the case for self-government is not made.[7]

[7] It should be made clear that these observations relate to the right to self-determination as it is commonly understood in the discourse of international relations and international morality. In principle, there could be a different right of self-determination, i.e., a right that, when the case for self-government is established, self-government should be granted, i.e., that all the international agents have a duty to take what action is necessary to grant self-government to the encompassing group regarding which the case for self-government has been established. That is, there could in principle have been a substantive right to have self-government when it is

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The problem in conceding the existence of such a right is, of course, not the possibility that a group that would best be self-governing does not wish to be so. Given the strong subjectivist element in the instrumentalist argument, such reluctance to assume independence would suggest that the case for its being self-governing is much weakened. The problem is that the case for self-government is hedged by considerations of the interest of people other than members of the groups, and by the other interests of members of the groups, i.e., other than their interests as members of the groups. These include their fundamental individual interests which should be respected, e.g., by a group whose culture oppresses women or racial minorities. These considerations raise the question whether encompassing groups are the most suitable bodies to decide about the case for self-government. Can they be entrusted with the decision in a matter in which their group interests are in conflict with other interests of members of the group as well as with the interests of other people? At the very least this suggests that the right must be qualified and hedged to protect other interests.

More fundamental still is the question of how the right of self-determination fits within our general conception of democratic decision making. We are used to a two-level structure of argument concerning social issues, such as just taxation, the provision of public education, etc. First, we explore the principles that should govern the matter at issue. Second, we devise a form of democratic procedure for determining what shall be done. The first level answers the question 'what should be done?'. The second responds to the question 'who should decide?'.

On a simple majoritarian view, the issue of self-government seems to defy a democratic decision procedure. The question is 'what is the relevant democratic unit?' and that question cannot be democratically decided, at least not entirely so. In fact, of course, we are not simple majoritarians. We adopt a whole range of democratic procedures such as constitution-making privileged majorities, ordinary legislative processes, plebiscites, administrative processes, and decisions by special agencies under conditions of public accountability and indirect democratic control. We match various democratic processes with various social and political problems. This means that there is no universal democratic formula serving as the universal

---

right that one should have it, rather than a "who is to decide" right, that an encompassing group should be entitled to decide whether it should be self-governing. Below we touch briefly on the reasons that explain why the right of self-determination as we know it today is not of this kind.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

answer to 'who decides?' questions. Rather, we operate a mixed principled-democratic system in which principles, whose credentials do not derive entirely from their democratic backing, determine what form of a democratic procedure is suited for what problem. Within this mixed principled-democratic framework, the right to self-determination fits as just another qualified democratic process suited to its object.

What are the principles involved? It is tempting to see here a principle giving the part veto over the issue of membership in a larger whole. To form a new political unit, or to remain part of an existing one, all component parts should agree. To break up a political unit, or to foil the creation of a new one, all that is required is the will of the group that wants to secede or to stay out. This principle derives its appeal from its voluntaristic aura. It seems to regard the justification of all political units as based on consent. But this is an undesirable illusion. It is undesirable since, as was explained above regarding encompassing groups, the more important human groupings need to be based on shared history, and on criteria of nonvoluntaristic (or at least not wholly contractarian) membership to have the value that they have. The principle presents no more than an illusion of a contractarian principle since it refers to groups, not to individuals. But the whole contractarian ethos derives its appeal from the claim that each individual's consent is a condition of the legitimacy of political units. Beyond all that, the principle simply begs the question that it is meant to answer, namely, what are the parts? Which groupings have the veto and which do not? Can the group of all the people whose surnames begin with a '*g*' and end with an '*e*' count for these purposes? Do they have the veto on membership in a larger political unit?

The right to self-determination derives from the value of membership in encompassing groups. It is a group right, deriving from the value of a collective good, and as such opposed in spirit to contractarian-individualistic[8] approaches to politics or to individual

[8] The reference is to moral individualism, or value individualism, not to methodological individualism. It is impossible here to deal with the matter adequately. Let us simply indicate our position briefly. There is no accepted characterization of the term. In *The Morality of Freedom* [(New York: Oxford, 1986), p. 198], Raz identified moral individualism with the view that only individual goods, and no collective goods, have intrinsic values. According to individualism so understood, membership of encompassing groups, and the prosperity of such groups, cannot be of intrinsic value. But we believe that it is intrinsically valuable. Hence, on this definition our approach is not individualistic. In "Three Grades of Social Involvement" [*Philosophy and Public Affairs*, XVIII (1989), p. 133], George Sher characterizes

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

well-being. It rests on an appreciation of the great importance that membership in and identification with encompassing groups has in the life of individuals, and the importance of the prosperity and self-respect of such groups to the well-being of their members. That importance makes it reasonable to let the encompassing group that forms a substantial majority in a territory have the right to determine whether that territory shall form an independent state in order to protect the culture and self-respect of the group, provided that the new state is likely to respect the fundamental interests of its inhabitants, and provided that measures are adopted to prevent its creation from gravely damaging the just interests of other countries. This statement of the argument for the right requires elaboration.

(1) The argument is an instrumental one. It says, essentially, that members of a group are best placed to judge whether their group's prosperity will be jeopardized if it does not enjoy political independence. It is in keeping with the view that, even though participation in politics may have intrinsic value to individuals, the shape and boundaries of political units are to be determined by their service to individual well-being, i.e., by their instrumental value. In our world, encompassing groups that do not enjoy self-government are not infrequently persecuted, despised, or neglected. Given the importance of their prosperity and self-respect to the well-being of their members, it seems reasonable to entrust their members with the right to determine whether the groups should be self-governing. They may sacrifice their economic or other interests for the sake of group self-respect and prosperity. But such a sacrifice is, given the circumstances of this world, often not unreasonable.

One may ask why should such matters not be entrusted to international adjudication by an international court, or some other international agency. Instead of groups' having a right to self-determination which makes them judges in their own cause, the case for a group's

---

moral individualism as the belief that moral justification proceeds through premises relating to individuals and their preferences. His characterization is too vague to be conclusively disputed (e.g., all holistic justifications will include premises relating to preferences as well as to everything else—does that make them individualistic?). But if Sher has in mind the standard type of (actual or hypothetical) contractarian justifications, then our approach is not individualistic. Because actual individual preferences heavily depend on social practices, there is no reason to give them justificatory primacy. The content of hypothetical preferences is either too indefinite to yield any results or is made definite by assuming a certain social context to give them meaning. Either way it cannot be endowed with justificatory primacy, though of course people's capacity to respond to various conditions, and to form various goals and attachments, is central to any moral justification.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

becoming self-governing should be entrusted to the judgment of an impartial tribunal. This would have been a far superior solution to the question 'who is to decide?'. Unfortunately, there simply does not exist any international machinery of enforcement that can be relied upon in preference to a right of self-determination as the right of self-help, nor is there any prospect of one coming into existence in the near future. In the present structure of international relations, the most promising arrangement is one that recognizes group rights to self-determination and entrusts international bodies with the duty to help bring about its realization, and to see to it that the limits and preconditions of the right are observed (these are enumerated in the points two to five below).

(2) The right belongs to the group. But how should it be exercised? Not necessarily by a simple majority vote. Given the long-term and irreversible nature of the decision (remember that while independence is up to the group, merger or union is not), the wish for a state must be shared by an overwhelming majority, reflecting deep-seated beliefs and feelings of an enduring nature, and not mere temporary popularity. The precise institutional requirements for the exercise of the right are issues that transcend the topic of this paper. They are liable to vary with the circumstances of different national and ethnic groups. Whatever they are they should reflect the above principle.

(3) The right is over a territory. This simply reflects the territorial organization of our political world. The requirement that the group be a substantial majority of the territory stems from further considerations aimed at balancing the interest in self-government against the interests of nonmembers. First, it is designed to ensure that self-government for a territory does not generate a problem as great as it is meant to solve, by ensuring that the independence will not generate a large-scale new minority problem. That risk cannot be altogether avoided. As was remarked before, numbers count in the end.

A further factual assumption underlying this condition is that people are, even today, most directly affected by the goings-on in their region. It is true that one's economic conditions are affected by the economic activities in far away places. This, however, is more and more true of the international system generally. The ideal of economic autarchy died a natural death. (Correspondingly, the condition of economic viability which used to figure in theories of the states in international relations has little role in the modern world.) What can be secured and protected, and what vitally matters to the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

quality of life, is its texture as determined by the local culture and custom, the nature of the physical environment, etc. Hence the right is given only to a group that is the majority in a territory. The case for self-government applies to groups that are not in the majority anywhere, but they do not have the right to self-determination anywhere. Their members, like other people, may have a right to immigration on an individual basis to a territory of their choice. But their case is governed by general principles of freedom of movement and the sovereign rights of existing states. This means that their communal interests remain an important consideration to be born in mind by the decision makers, but they have no right, i.e., the decision is not up to them.

Do historical ties make a difference? Not to the right if voluntarily abandoned. Suppose that the group was unjustly removed from the country. In that case, the general principle of restitution applies, and the group has a right to self-determination and control over the territory it was expelled from, subject to the general principle of prescription. Prescription protects the interests of the current inhabitants. It is based on several deep-seated concerns. It is meant to prevent the revival of abandoned claims, and to protect those who are not personally to blame from having their life unsettled by claims of ancient wrongs, on the ground that their case now is as good as that of the wronged people or their descendants. Prescription, therefore, may lose the expelled group the right even though its members continue to suffer the effects of the past wrong. Their interest is a consideration to be borne in mind in decisions concerning immigration policies, and the like, but because of prescription they lost the right to self-determination. The outcome is not up to them to decide.

(4) The right is conditional on its being exercised for the right reasons, i.e., to secure conditions necessary for the prosperity and self-respect of the group. This is a major protection against abuse. Katanga cannot claim a right to self-determination as a way of securing its exclusive control over uranium mines within its territory. This condition does not negate the nature of a right. The group is still entrusted with the right to decide, and its decision is binding even if wrong, even if the case for self-government does not obtain, provided the reasons that motivate the group's decision are of the right kind.

(5) Finally, there are the two broad safeguards on which the exercise of the right is conditional. First, that the group is likely to respect the basic rights of its inhabitants, so that its establishment will do good rather than add to the ills of this world. Secondly, since the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

establishment of the new state may fundamentally endanger the interests of inhabitants of other countries, its exercise is conditional on measures being taken to prevent or minimize the occurrence of substantial damage of this kind. Such measures, which will vary greatly from case to case, include free-trade agreements, port facilities, granting of air routes, demilitarization of certain regions, etc.

Two kinds of interests do not call for special protection. One is the interest of a people to regard themselves as part of a larger rather than a smaller grouping or country. The English may have an interest in being part of Great Britain, rather than mere Englanders. But that interest can be justly satisfied only with the willing co-operation of, e.g., the Scots. If the other conditions for Scottish independence are met, this interest of the English should not stand in its way. Secondly, unjust economic gains, the product of colonial or other form of exploitation of one group by another, may be denied to the exploiting group without hesitation or compensation (barring arrangements for a transitory period). But where secession and independence will gravely affect other and legitimate interests of other countries, such interests should be protected by creating free-trade zones, demilitarized areas, etc.

(6) A right in one person is sufficient ground to hold some other person(s) to be under a duty.[9] What duties arise out of the right to self-determination? How is this matter to be settled? As the previous discussion makes clear, the right of self-determination is instrumentally justified, as the method of implementing the case for self-government, which itself is based on the fact that in many circumstances self-government is necessary for the prosperity and dignity of encompassing groups. Hence, in fixing the limits of the right, one has to bear in mind the existing system of international politics, and show that, given other elements in that system, certain duties can be derived from the right to self-determination, whereas others cannot. The first and most important duty arising out of the right is the duty not to impede the exercise of the right, i.e., not to impede groups in their attempts to decide whether appropriate territories should be independent, so long as they do so within the limits of the right. This duty affects in practice first and foremost the state that governs the territory concerned and its inhabitants.

There may be other duties following from the right of self-deter-

[9] See Raz, *The Morality of Freedom*, ch. 7: The Nature of Rights. On the relations of moral and legal rights, see also Raz, "Legal Rights," *Oxford Journal of Legal Studies*, IV (1984), p. 1. Raz has applied this analysis to the case of constitutional rights in general in ch. 10 of *The Morality of Freedom*.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

mination. In particular, there may be a duty on the state governing the territory to provide aid in exercising the right, and a duty on other states to aid the relevant group in realizing its right, and thus to oppose the state governing the territory if it impedes its implementation. But the extent of these duties must be subject to the general principles of international morality, which indicate what methods may and may not be used in pursuit of worthwhile goals and in preventing the violation of rights. As indicated at the outset, the examination of the details of such implications of the right is beyond the scope of this article.

This brings to an end our consideration of the outlines of the case for a right to self-determination and its limits. It is an argument that proceeds in several stages from fundamental moral concerns to the ways in which they can be best implemented, given the way our world is organized. The argument is meant to present the normal justification for the right. It does not claim that there could not be alternative justifications. But it does claim to be the central case, which alternatives presuppose or of which they are variations.[10]

Two conclusions emerge from this discussion. On the one hand, the right to self-determination is neither absolute nor unconditional. It affects important and diverse interests of many people, from those who will be citizens of the new state, if it comes into being, to others far away from it. Those who may benefit from self-government cannot insist on it at all costs. Their interests have to be considered along those of others. On the other hand, the interests of members of an encompassing group in the self-respect and prosperity of the group are among the most vital human interests. Given their importance, their satisfaction is justified even at a considerable cost to other interests. Furthermore, given the absence of effective enforcement machinery in the international arena, the interest in group prosperity justifies entrusting the decision concerning self-government to the hands of an encompassing group that constitutes the vast majority of the population in the relevant territory, provided other vital interests are protected.

AVISHAI MARGALIT
The Hebrew University of Jerusalem

JOSEPH RAZ
Oxford University

[10] On the notion of a "normal justification," and the reasons why it cannot be analyzed as either a necessary or a sufficient condition, see Raz, *The Morality of Freedom*, ch. 3.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

TROUBLES WITH OCKHAMISM*

SUPPOSE that Jack signs his work contract on January 3, 1985. According to many traditional theologians, there is no contradiction between Jack's *freely* performing this act on January 3 and God's foreknowing this fact, say, on January 1.[1] What follows logically from the assumption that

(K) God foreknows on January 1, 1985 that Jack will sign his contract on January 3, 1985.

is only that Jack, in fact, will do so, but not that he has to or must do so. On this account, God not only knows that Jack will sign the contract on the said date, but he also knows that Jack will do so freely, i.e., that it is in his power to do otherwise. In the mid-1960s, Nelson Pike launched a vigorous attack on this common view in his well-known article, "Divine Omniscience and Voluntary Action,"[2] presenting an argument purporting to show that divine foreknowledge and human freedom are inconsistent. Here is a version of Pike's argument.

(TF)	(a)	'God' is a proper name. 't_1', 't_2', etc., refer respectively to January 1, 1985; January 2, 1985; etc. (Premise)
	(b)	God exists at every moment of time, and is essentially omniscient in the sense that it is logically impossible for him to believe a false proposition, or to fail to believe any true proposition. (Premise)
	(c)	Jack signs his contract at t_3, and it is within Jack's power at t_2 to bring it about that he does not sign his contract at t_3. (Premise)
	(G)	God believes at t_1 that Jack signs his contract at t_3. [By (b) and (c)]
	(PRW)	If q is a logically necessary condition for p, and q does not obtain, then an agent has it within his power at t to bring it about that p only if he has it within his power at t to bring it about that q. (Here p and q range over dated states of affairs like the sun is shining at t_3, Jack eats lunch at t_4, etc.) (Premise)[3]

* I am greatly indebted to John Fischer, Harry Friedman, Jerome Gellman, Dale Gottlieb, Charlotte Katzoff, Philip Quinn, William Rowe, and the editors of this JOURNAL for excellent comments on an earlier version of this paper.

[1] The term 'freely performing' I understand here in the libertarian sense.

[2] *The Philosophical Review*, LXXIV (1965): 31–35.

[3] This principle is an improved version of a principle employed by Richard Taylor

0022-362X/90/8709/462–480 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(d) That it is not the case that God believes at t_1 that Jack signs his contract at t_3, is a logically necessary condition for Jack's not signing his contract at t_3. [By (b)][4]

(e) It is within Jack's power at t_2 to bring it about that God does not believe at t_1 that Jack signs his contract at t_3. [By (c), (d), (G) and (PRW)] [5]

But, (e) contradicts the intuitive principle that we do not have power over the past. We may refer to this principle as *the principle of the fixity of the past* or PFP.

One way of responding to this argument is Ockhamism. The Ockhamist distinguishes between two sorts of facts about the past; genuine or *hard facts* about the past which are subject to the principle of the fixity of the past; and *soft facts* about the past which, since being somehow also about the future, do not intuitively fall under this principle. For example, facts such as

(1) Hitler attacked Russia in 1941.
(2) Smith uttered on January 1, 1985 the sentence 'Jack will sign his contract on January 3, 1985'.
(3) Smith believed on January 1, 1985 that Jack will sign his contract on January 3, 1985.

---

in "Fatalism," *The Philosophical Review*, LXXI (1962), p. 58. Cf. William L. Rowe, "Fatalism and Truth," *The Southern Journal of Philosophy*, XVIII (1980), p. 217. Note that (PRW) ought to be distinguished from the following fallacious principle:

(PR) If q is a logically necessary condition for p, then an agent has it within his power at t to bring it about that p only if he has it within his power at t to bring it about that q.

The fact that q is a necessary condition for p does not imply that my having the power to bring about p is contingent upon my having the power to bring it about that q. I may have the power to bring about p, even if I lack the power to bring about q in situations where q is already in existence or will be in existence anyway. E.g., provided I did not smoke a cigarette until t_5, I may have the power at t_4 to smoke a cigarette for the first time at t_5, and this even though I do not have at t_4 the power to bring it about that I did not smoke a cigarette in the past. Or given that the sun will be shining at t_5, I may have the power at t_2 to bring it about that I raise my arm at t_3 while the sun is shining. But surely, I do not have the power at t_2 to bring it about that the sun will be shining at t_3. For similar counterexamples to (PR), see Joshua Hoffman and Gary Rosenkrantz, "On Divine Foreknowledge and Human Freedom," *Philosophical Studies*, XXXVII (1980), pp. 292–3. I discuss the relation between (PR) and (PRW) also in my "Two Forms of Fatalism," in John Fischer, ed., *God, Freedom and Knowledge* (Stanford: University Press, 1989).

[4] Here, as well as in what follows, I shall use terms such as 'necessarily' and 'logically necessary' to signify broadly logical necessity.

[5] This version of Pike's argument differs somewhat from Pike's original formulation of it (see pp. 33–5). It has the advantage of spelling out precisely the transition from (i) Jack's having it within his power at t_2 to bring it about that it is not the case that Qt_3, to (ii) his having the power at t_2 to bring it about that God did not believe at t_1 that Qt_3. (Qt_3 abbreviates 'Jack signs his contract on January 3, 1985'.)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

are considered hard facts about the past, relative to January 2, 1985, whereas

(4) Hitler attacked Russia in 1941, 44 years before Jack's signing his contract on January 3, 1985.
(5) Smith uttered on January 1, 1985 the *true* sentence 'Jack will sign his contract on January 3, 1985'.
(6) Smith correctly believed on January 1, 1985 that Jack will sign his contract on January 3, 1985.

are paradigm examples of soft facts about the past, relative to January 2, 1985. Furthermore, the Ockhamist claims that facts about God's foreknowledge of future events, and more specifically a fact such as

(G) God believes at t_1 that Jack will sign his contract at t_3.[6]

may be treated as soft facts about the past on a par with (4)–(6). Consequently, since the power attributed to Jack in (e) is the power to bring about the nonobtaining of a soft fact about the past, the principle of the fixity of the past is not violated.[7]

In this paper, I wish to question the adequacy of the Ockhamistic approach to reconciling divine foreknowledge with human freedom.[8] I argue that there is an important asymmetry between (G) and the likes of (4)–(6) which bears importantly on the question of their fixity. The attempt to treat (G) as a soft fact about the past on a par with (4)–(6) is in my view misguided. The discussion to follow is divided into four parts. In part I, I examine the reasons offered by some contemporary Ockhamists for treating (G) as a soft fact about the past, and argue that these reasons are either of doubtful relevance or rest on assumptions that are mistaken.[9] In parts II and III, I

[6] Note that, in view of the fact that God is essentially omniscient, (G) is logically equivalent with the fact (K) God foreknew at t_1 that Jack will sign his contract at t_3.

[7] This solution to the problem of divine foreknowledge and human freedom bears a close similarity to a solution to the problem suggested by the fourteenth-century philosopher William of Ockham. For an excellent presentation of Ockham's view, see Marilyn Adams and Norman Kretzmann's introduction to their translation of Ockham's *Predestination, God's Foreknowledge and Future Contingents* (New York: Appleton Century Crofts, 1969).

[8] Thus, the optimism that I expressed in an earlier article with regard to the prospects of working out an acceptable Ockhamistic position seems to me now to have been a bit premature. See Eddy M. Zemach and David Widerker, "Facts, Freedom and Foreknowledge," *Religious Studies*, XXIII (1987): 19–28.

[9] Contemporary Ockhamistic accounts have been also criticized by John Fischer in "Freedom and Foreknowledge," *The Philosophical Review*, XCII (1983): 69–79; "Ockhamism," The *Philosophical Review*, XCIV (1985): 81–100; and "Hard-Type Soft Facts," *The Philosophical Review*, XCV (1986): 591–601; and by William Hasker in "Hard Facts and Theological Fatalism," *Noûs*, XXII (1988): 419–436.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

offer an account of the intuitive nonfixity of facts like (4)–(6), and argue that, on pain of leading to counterintuitive consequences, this account cannot be applied to facts about divine foreknowledge. Finally, in part IV, I examine a recent attempt by John Fischer which is also intended to bring out the asymmetry between (G) and standard cases of soft facts and contend that it is incomplete.

I

There are two requirements that must be met by the Ockhamist in order for him to be justified in treating (G) as a soft fact about the past on a par with (4)–(6).

R1. He must specify a property that (G) shares with (4)–(6).
R2. Most importantly, he must give us a good reason to think that it is *in virtue* of having this property or feature that facts like (4)–(6) are not subject to PFP.

The idea underlying this last requirement is clear. It is intended to rule out attempts to assimilate (G) to (4)–(6) based on similarities between these two sorts of facts which are irrelevant to the issue of past fixity. Let us examine now whether contemporary Ockhamistic accounts can be said to meet these requirements.

In describing the Ockhamistic position, William Rowe,[10] for example, claims that the softness of (4)–(6) is due to the fact that they entail a fact about the future. Indeed, this is the reason why, according to him, such facts are not subject to PFP. He refers to such facts as "facts which are *not simply* about the past," as opposed to those which are "*simply* about the past." Consequently, since (G) shares this feature with (4)–(6), it also is not subject to PFP. He says:

> . . . our conviction that the past is beyond our power to affect is certainly true, so far as facts which are simply about the past are concerned. Facts which are about the past, but *not simply* about the past, may not, however be beyond our power to affect. And what Ockham saw is that facts about divine foreknowledge which are used as the basis for denying human freedom are facts about the past, but *not simply* about the past (*ibid.*, p. 164).

But what has a past fact's entailing a fact about the future got to do with its not being subject to PFP? What is the conceptual link between a past fact's entailing a fact about the future and its being immune to this principle? We may agree with Rowe that (4)–(6) do not seem subject to PFP and that they entail a fact about the future. But is it *because* of having this feature that PFP does not apply to them? On this issue, Rowe leaves us completely in the dark. Also, the

[10] *Philosophy of Religion* (Encino: Dickenson, 1978), pp. 162–5.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

mere fact that a past fact entails a fact about the future does not render it immune to PFP. For example, the fact that Jack raises his arm at t_1, entails the future fact that it is not the case that he raises his arm for the first time at t_3. But despite this entailment, once having raised his arm at t_1, Jack no longer has it within his power not to raise his arm at t_1.[11]

Consider next the Ockhamistic position as expounded by Marilyn Adams[12] and refined later by Joshua Hoffman and Gary Rosenkrantz.[13] Adams's idea is that a hard fact about a past time t, i.e., the kind of past fact to which PFP does apply, is a fact which is at least in part about t, but is not about any time that is future relative to t. (Or stating it more precisely, it is a fact that may be expressed by a statement that is at least in part about t, but is not about any time that is future relative to t.) She then provides a definition of 'hard fact about t', one central component of which is that such a fact (or the statement expressing it) must not entail the obtaining of a state of affairs at a time later than t.[14] On this account, (4)–(6) as well as (G)

[11] See Fischer, "Hard-Type Soft Facts," p. 593.

[12] "Is the Existence of God a 'Hard Fact'," *The Philosophical Review*, LXXVI (1967), pp. 493–4.

[13] "Hard and Soft Facts," *The Philosophical Review*, XCIII (1984): 419–434.

[14] According to Adams,

(B) "Statement p is at least in part about time T" = df. "The happening or not happening, actuality or non-actuality of something at T, is a necessary condition of the truth of p."

(C) "Statement p expresses a hard fact about a time T" = df. "p is not at least in part about any time future to T" (pp. 492–3).

Hoffman and Rosenkrantz give a much more elaborate account of the notion of a hard fact about a time t. What is relevant for our purpose here, however, is that they, too, require that a hard fact about t must not entail the obtaining of an "immediate" state of affairs or what they call an "URP" state of affairs at a time later than t. By the latter they understand such simple state of affairs, such as Socrates is walking, Socrates is wise, Jack mows his lawn, etc. (pp. 425–6). My criticism of Adams's position is not affected by this refinement.

Notice that Adams's account is not vulnerable to the objection that since

(s1) Jack raises his arm t_1.

entails

(s2) It is not the case that Jack raises his arm for the first time at t_3.

(s1) is in part about t_3, and hence the fact expressed by it cannot be deemed a hard fact about t_1. (See Fischer, "Hard-Type Soft Facts," p. 593, and "Freedom and Foreknowledge," p. 75.) Rebuttal: Although (s1) entails (s2), (s1) does not entail, as is required by Adams's account, that

(i) Some state of affairs obtains *at* t_3.

nor, more specifically, that

(j) The state of affairs that it is not the case that Jack raises his arm for the first time at t_3, obtains *at* t_3.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

turn out to be a soft facts about t_1, since they entail that the state of affairs that Jack signs the contract obtains at t_3. As in the case of Rowe, we may ask: What bearing does a past fact's entailing the obtaining of a state of affairs in the future have upon its not being subject to PFP? To be sure, (G) differs from a hard fact such as (1) in that it entails the obtaining of a state of affairs at t_3. But why should this difference be a relevant one? We may gain a better understanding of Adams's position by inquiring what led her to try to capture the notion of a hard fact about a past time t the way she did in the first place. She says:

> It is useful before proceeding . . . to examine the distinction between "hard" and "soft" facts. Pike makes the distinction between "hard" and "soft" facts about the past by contrasting facts which were "fully accomplished" or "over-and-done with" at a given past time with those which were not. I think that the distinction Pike has in mind can also be drawn in terms of a statement's being about a given time (*op. cit.*, p. 493).

Adams, as we can see, takes the notion of a hard fact about a time t to be equivalent in meaning or closely related to the notion of a fact's being fully accomplished or over-and-done-with at times after t. This by itself is perfectly acceptable, since we can agree that a fact that has this feature is fixed at times after t. But by requiring that a hard fact about t should not entail the obtaining of a certain state of affairs at a time later than t, Adams makes the further assumption that

> (HD) A fact about t that entails the obtaining of a state of affairs at a time t' later than t cannot be deemed over-and-done-with before t'.[15]

This assumption is problematic. To see this, let us suppose that God promises Smith at t_1 that Jack will sign the contract at t_3. Suppose further that it is impossible for God to break his promises, i.e.,

> (I) Necessarily, if God promises to Smith at t_1 that Qt_3, then Qt_3.

(Qt_3 abbreviates 'Jack signs his contract on January 3, 1985'.) It seems intuitive to assume that the fact that God promised Smith at t_1

For instance, (i) are (j) are both false in a possible world in which John raises his arm at t_1, but in which there are no times later than t_1. But in such a world, (s1) and (s2) would still be true. On this point, see also my "Two Fallacious Objections to Adams's Soft/Hard Fact Distinction," *Philosophical Studies* (forthcoming).

[15] Or stating this assumption alternatively as follows:

(HD) A fact about the past, relative to a time t', which entails the obtaining of a state of affairs at a time later than t', cannot be deemed over-and-done-with at t'.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that Jack will sign the contract at t_3 is fully accomplished and over-and-done-with at t_2. But, on (HD), this fact would not have this feature, since it entails a fact about t_3. To take another example, suppose that God decrees at t that of those (over twenty years old) who rebelled against Moses after the return of the spies, only Caleb and Joshua will enter the land of Canaan at t' ($t' > t$). Abbreviate the statement expressing the decree to Rt'. Suppose further that God's will cannot be thwarted, i.e.,

(J) Necessarily, if God decrees at t that P, then P.

Again, the fact that God decreed at t that Rt' seems intuitively to be over-and-done-with immediately after t, although it entails the obtaining of a state of affairs at a time later than t. The moral to be drawn from these examples is that an assumption such as (HD) is justified as long as we do not allow for the possibility of there being logically (or metaphysically) necessary connections between temporally distinct events.[16] Once we deviate from this policy, however, and allow for such necessary connections between events as those which obtain between divine forebeliefs and human free actions, we can easily conceive of situations in which (HD) does not hold.

Another philosopher to take the Ockhamistic line is Alvin Plantinga.[17] His justification for treating (G) as a soft fact about the past is the following. Plantinga starts from the assumption that

(PL1) A hard fact about the past must be one that is *strictly* about the past.[18]

Although Plantinga does not offer us an account of strictness in this sense, he gives us a sufficient condition for nonstrictness, which is that

(PL2) No conjunctive proposition that contains as a conjunct a simple proposition about the future, like Paul will mow his lawn in 1999, can be deemed strictly about the past.

So, for example, the conjunctive proposition

[16] Here, as well as in what follows, I take 'event' to refer to concrete, dated, nonrepeatable events. Also, I include under the rubric 'event' states of substances and processes. In addition to events, I countenance also such entities as states of affairs and facts.

[17] "On Ockham's Way Out," *Faith and Philosophy*, III (1986), pp. 247–8. See especially the passage starting with: "We may not be able to give a criterion for being strictly about the past . . ." (p. 248).

[18] Plantinga also says that a hard fact about the past must be a *genuine* fact about the past (*ibid.*, p. 247). Aside from making only some cursory remarks about what it is for a fact to be a genuine fact about the past, however, he does not explain this notion.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(7) Smith believed at t_1 that Jack will sign the contract at t_3, and Jack will sign the contract at t_3.

would not be strictly about the past, relative to t_2, since it contains as a conjunct the proposition

(8) Jack will sign the contract at t_3.

Plantinga's third assumption is that

(PL3) Hard facthood is closed under logical equivalence.

Consequently, since (G) is logically equivalent to the conjunction of (G) and (8), (G) is not a hard fact about the past, according to Plantinga. This argument for the soft facthood of (G) is again unconvincing. The problematic assumption is (PL3). As in the case of (HD), once we allow for the possibility of logically (or metaphysically) necessary connections between temporally distinct events, there seems to be no reason to accept (PL3). The divine promise example mentioned earlier is revealing here as well. The proposition

(9) God promises to Smith at t_1 that Jack will sign the contract at t_3.

surely expresses a hard fact about the past (relative to t_2), in the sense of that fact's being over-and-done-with at t_2. It expresses a fact that at t_2 can be remembered by Smith, can be empirically detected by him, e.g., by hearing the promise, and may have traces at t_2 just like regular hard facts about the past (relative to t_2) do. According to Plantinga, however, (9) would not express a hard fact about the past; for it is logically equivalent to the conjunction

(10) God promises to Smith at t_1 that Jack will sign the contract at t_3, and Jack will sign the contract at t_3.[19]

What the above discussion suggests is that the reasons offered by contemporary Ockhamists for treating (G) as a soft fact about the past on a par with (4)–(6) are either of doubtful relevance to the freedom and foreknowledge debate, or else are based upon assumptions that are very problematic. In effect, these philosophers have failed to give us a plausible explanation why soft facts such as (4)–(6) do not seem to be subject to PFP. The question that arises now is whether someone who is attracted to the Ockhamistic strategy of

[19] To be sure, Plantinga might argue that he is operating with a notion of a hard fact about the past according to which the fact expressed by (9) would not count as a hard fact about t_1. But then he still owes us an explanation why this fact, despite its being intuitively over-and-done-with at t_2, and despite its having all the other features mentioned in the text, is not subject to PFP.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

reconciling divine foreknowledge with human freedom can do better?

II

What is it that accounts for the intuitive nonfixity of facts like (4)–(6)? To answer this question, let us consider more closely the following version of the principle of the fixity of the past:

(PFP1) If a given event occurs at a time t, then no one has it within his power at a time later than t to bring it about that that event did not occur at t.

or more formally put:

$$O(e, t) \supset \sim(\exists X)(\exists t')[t' > t \cdot P^*_{X,t'}(O(e, t))]$$

(where e, t, and X range respectively over events, times, and agents, and O and $P^*_{X,t'}$ are short for 'occurred' and 'it is within X's power at t' to bring it about that').

Notice that, although we do not have the power to bring about the nonoccurrence of a given past event E, we sometimes have the power to bring it about that that event does, or does not exemplify certain future-contingent properties, namely, those properties it exemplifies *in virtue* of standing in certain relations to events or states of affairs occurring in the future. So, for example, Jack does not have it within his power, on January 2, 1985, to bring about the nonoccurrence of events such as

E1: Smith's uttering on January 1 the sentence 'Jack will sign the contract on January 3'.
E2: Smith's believing on January 1 that Jack will sign the contract on January 3.

Given that Jack is free with respect to signing the contract, however, it is within his power on January 2 to bring it about that E1 and E2 do or do not exemplify properties like

P1: occurring two days before Jack's signing the contract on January 3
P2: being such that Jack signs the contract on January 3
P3: being an uttering of the true (false) sentence 'Jack will sign the contract on January 3'

Once having drawn the distinction between

(i) having the power to bring about the nonoccurrence of a certain past event

and

(ii) having the power to bring it about that a past event does or does not exemplify a certain future contingent property

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

we are in a position to provide an adequate answer to the question why soft facts like (4)–(6) are intuitively not fixed. The answer is this. Each of these facts may be viewed as an exemplification of a future-contingent property P by some past event E. Hence, to the extent that an agent has the power to bring about the nonoccurrence of the future event or fact in virtue of which E exemplifies P, that agent also has the power to bring about the nonobtaining of the relevant soft fact. By crediting him with that power, we do not credit him with the power to bring about the nonoccurrence of a past event, but merely with the innocuous power to make it the case that a past event does not exemplify a certain future-contingent property. To be sure, an agent does not always have the power to bring about the nonoccurrence of the future event that figures in a soft fact. Sometimes this may be due to the fact that the agent does not have the knowledge or the means to prevent the occurrence of the event. In other cases, he may lack that power because the future event is causally determined, as, e.g., in the case of the sun's shining at t_3. Thus, the following facts, though being exemplifications of some future-contingent property by an event occurring at t_1 (and hence soft facts about t_1), would still be fixed for Smith, relative to t_2.

(11) Smith saw Jack at t_1, two days before the event of his mother's getting cancer at t_3.

(12) Smith correctly believed at t_1 that the sun will be shining at t_3.

But their fixity (relative to t_2), as opposed to the fixity of a hard fact about t_1 like

(13) Smith saw Jack at t_1.

does not derive from their being fully accomplished and over-and-done-with at t_2, i.e., it does not derive from PFP.[20]

It may be useful at this point to state the idea of a future-contingent property of an event schematically as follows:

(FCP) P is a future-contingent property of an event E_t, occurring at a time t, iff

1. E_t exemplifies P
2. The state of affairs of E_t exemplifying P can be analyzed either
   (i) as a relation of some sort obtaining between E_t and some event $E_{t'}$, which occurs at a time t' later than t, i.e., $R(E_t, E_{t'})$, $t' > t$,

   or it can be analyzed

[20] Cf. Rowe, *op. cit.*, p. 165; and Fischer, "Hard-Type Soft Facts," p. 595.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(ii) as a relation of some sort obtaining between Et and some fact Cf that is future-based, relative to t, i.e., $R(Et, Cf)$. A future-based fact, relative to t, will typically be a fact expressible by a true proposition Pt' to the effect that an object exemplifies a certain property at a time t' later than t. Or it will be a fact expressible by a proposition R that is 1-entailed by some such proposition Pt'. (The term '1-entails' applies here only to first-order logic entailments.)[21]

So, for example, $P1$ is a future-contingent property of $E1$, since the state of affairs of $E1$'s exemplifying $P1$ can be analyzed as a relation obtaining between the event of Smith's uttering at t_1 the sentence 'Jack will sign his contract on January 3, 1985', and the event of Jack's signing the contract on January 3. Similarly, $P2$ is a future-contingent property of $E2$, since the state of affairs of $E2$ exemplifying $P2$, can be analyzed as a relation holding between $E2$ and the fact that Jack signs the contract at t_3, which is future-based, relative to t_1, i.e., $E2$'s being such that Jack signs the contract at t_3.

III

Given the above account of the intuitive nonfixity of soft facts such as (4)–(6), let us ask now whether the Ockhamist can use this account to justify his claim regarding the alleged nonfixity of (G). In other words, can the Ockhamist offer a relational analysis of (G) on the pattern of $R(Et, Et')$ or $R(Et, Cf)$, and then claim that the assumption as to Jack's power at t_2 to bring about the nonobtaining of (G) does not imply that Jack has the power to bring about the nonoccurrence of a past event, but merely implies that he has the power to bring it about that some past event does not stand in some relation to some future event or fact? To answer this question, let us first ask: What would be the elements figuring in such a relation R? One possibility that immediately suggests itself is to identify these elements with the event

EG: God's believing at t_1 that Qt_3

and the fact

Qt_3: Jack signs the contract at t_3.

[21] These are only some informal, explicatory remarks on the notion of a state of affairs which is future-based, relative to t. I do not attempt in this paper to give a precise account of this notion. The discussion to follow is not affected by this limitation. Another way of informally explicating the notion of a future-based fact is to characterize it as a fact expressible by a true proposition R *made true* by the occurrence of some future event.

This schematic account of the notion of a future-contingent property of an *event* occurring at t can be easily generalized to yield a similar account of the notion of a future-contingent property of an *object* existing at t.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

That is, the suggestion here is to view (G) as some sort of relational fact

(G*) $R(EG, Qt_3)$

This suggestion, however, would be of no help to the Ockhamist. The Ockhamist wants to claim that it is within Jack's power to bring about the nonobtaining of (G), i.e., (G*). Note, however, that, in order to be credited with the power to bring about the nonobtaining of (G*), Jack must have the power to bring about the nonobtaining of Qt_3. But since the nonobtaining of Qt_3 entails the nonoccurrence of EG, it follows [given (PRW) and given the fact that EG occurred] that Jack can be said to have that latter power only if he has the power to bring about the nonoccurrence of EG, i.e., only at the expense of violating PFP1. It is instructive to note that this objection would not apply in the case of a typical soft fact, such as

(6) Smith correctly believed at t_1 that Qt_3.

Although (6), too, can be analyzed as $R(E2, Qt_3)$, the nonobtaining of Qt_3 does not entail the nonoccurrence of $E2$. If Jack had not signed the contract, the event $E2$ of Smith's believing at t_1 that Qt_3 would still have occurred, but would have merely lacked the property of being a believing of a true proposition.

At this point, the Ockhamist might try to construe (G) as a relation that obtains between some mental state M of God *other than* EG, and Jack's future action of signing the contract at t_3, i.e., $R(M, Et_3)$. Of course, the question that immediately arises is what would be the nature of this relation R. In the absence of an answer to this question, the proposal in question appears to be empty.[22] To help the Ockhamist, let us, for the sake of discussion, consider a different account of divine foreknowledge. Instead of viewing the latter on the model of belief, let us conceive of it on the model of immediate or direct awareness. That is, let us suppose that, just as humans are directly aware of their thoughts, sensations, and perhaps some abstract entities, God is directly aware of future-contingent events or facts.[23] On this account, $R(X, Et_3)$ would denote the relational property of being an awareness of Et_3, a property had by some appropriate mental state M of God. The logical form of 'God knows at t_1 that Qt_3' would be then given by:

[22] For an attempt to work out an account of divine belief which deals with this question, see Zemach and Widerker, p. 26. But this account seems to me now unsatisfactory in view of the objection of future-contingent causal chains which I raise later.

[23] For a similar approach to God's knowledge, see William L. Alston, "Does God Have Beliefs," *Religious Studies*, XXII (1986): 287–306.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

$(\exists M)$ [M occurs at t_1 and $R(M, Et_3)$]

This account seems to me the most promising proposal an Ockhamist can come up with in his attempt to treat facts about God's knowledge of future-contingent events as soft facts about the past on a par with (4)–(6). Because

(K) God knows at t_1 that Qt_3.

is construed as a relation obtaining between some mental state M of God and Jack's action of signing the contract, Jack, by having power over one of the relata of R, also has the power to bring about the nonobtaining of (K). Given the relational analysis of God's knowledge, this power does not imply a power to bring about the nonoccurrence of M, but merely implies a power to bring it about that M does not have the future-contingent property of being an awareness of Et_3.

Despite its initial attractiveness, however, this analysis of God's knowledge of future-contingent events turns out to be unsatisfactory. One problem with it is the intelligibility of the assumption that God can be directly aware *at* t_1 of an event that occurs at t_3. It would appear that (i) if someone is directly aware at t_1 of a certain event, then that event also occurs at t_1 (in which case the assumption that God is directly aware at t_1 of Et_3 would be contradictory).[24] This seems to be clearly so in the case of our knowledge of our own mental states, where, e.g., the event of my feeling pain and that of my being aware of my feeling pain are simultaneous. In response to this objection, an Ockhamist might want to reject (i) and insist that God, unlike humans, *can* be directly aware of events before they occur. Although this reply seems to me to be unconvincing, I shall not discuss it here any further.[25] Instead, I shall develop a general objection to any sort of Ockhamist attempt to reconcile divine fore-

[24] I am indebted to the editors of this JOURNAL for the emphasis of this point. Note that one can at t_1 be aware of an event in the sense of having at t_1 an inner mental representation of it. But such an awareness would not be a direct awareness of the event *itself*.

[25] Another objection one might raise against the account of divine foreknowledge under consideration is that (i) being an awareness of Et_3 is an essential property of a mental state that has it. Consequently, the Ockhamist's claim that it is within Jack's power at t_2 to bring it about that God does not know at t_1 that Qt_3 results not only in Jack's having at t_2 the power to bring it about that M lacks the said property, but it also results in Jack's having the power to bring about the nonoccurrence of M, and hence involves a violation of PFP1. An interesting way in which the Ockhamist might try to respond to the above objection, aside from squarely rejecting the claim that being an awareness of Et_3 is an essential property an event, is the following:

(a) The Ockhamist might first extend Jaegwon Kim's notion of event so as to apply also to events that have among their constitutive objects other events.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

knowledge with human freedom by treating facts about God's foreknowledge of future contingent events as soft facts about the past over which an agent may have power. I shall try to show that power over such facts sometimes gives an agent power over clear hard facts about the past, and hence leads to a violation of PFP1. This point, I claim, holds independently of the specific analysis of divine foreknowledge one adopts: the belief analysis or the direct-awareness analysis. The objection to follow will be formulated, for the sake of inclusiveness, in terms of God's knowledge. Depending upon the specific account of divine foreknowledge the Ockhamist adopts, however, it can be restated by replacing 'God knows . . .' with 'God believes . . .' or with 'God is directly aware of . . .'.

IV. THE OBJECTION FROM FUTURE-CONTINGENT CAUSAL CHAINS

(S) Suppose that God knows at t_1 that Jack will freely pull the trigger at t_5, with the intention of killing Smith. Suppose further that, wanting to save Smith, God reveals this fact to Smith at t_3. As a result, Smith by taking appropriate precautions is able to save his life. Now, the Ockhamist concedes that, by having the power to refrain from attempting to kill Smith, Jack also has at t_4 the power to make it the case that God did not know at t_1 that Jack would attempt to kill Smith at t_5.[26] On the other hand, God's knowing that Jack will at-

---

E.g., events such as Jack's seeing the accident at t_1, or Jack's accepting Jim's offer at t_1, etc. [As remembered, according to Kim an event is conceived of as an n-tuple of objects exemplifying an n-adic property at a given time t, i.e.,

$$[(X1, X2 \cdots Xn), Pn, t]$$

where the elements of the n-tuple are called "the constitutive objects of the event," the said property is called "the constitutive property of the event," and t is called the "constitutive time of an event." See Kim, "Causation, Nomic Subsumption and the Concept of an Event," this JOURNAL, LXX, 8 (April 26, 1973):217–236.]

(b) He might then restrict the application of PFP1 only to events that do *not* have among their constitutive objects events occurring after their constitutive time t. Genuine past events, an Ockhamist might insist, should not have such elements among their constitutive objects. On this account, PFP1 would apply to events such as Jack's mowing his lawn at t_1, or Jack's raising his arm at t_1, i.e.,

[(Jack, the lawn), ① mows ②, t_1]

[(Jack, his arm), ① raises ②, t_1]

But PFP1 would not apply to events such as God's being directly aware of Et_5, where Et_5 is the event of Jack's pulling the trigger at t_5, i.e.,

[(God, Et_5, ① is aware of ②, t_1]

or to Peter's predicting the occurrence of Et_5, i.e.,

[(Peter, Et_5), ① predicts ②, t_1]

[26] Note that this point holds both in case the Ockhamist adopts the belief model of divine foreknowledge (see the conclusion of Pike's argument) and in case he adopts the direct-awareness model of it.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tempt to kill Smith is a condition that in the circumstances causally contributes to the occurrence of the event of God's warning Smith at t_3, in the sense of being (in the circumstances) a causally necessary condition for it. Surely, if God had not known at t_1 that Jack would try to kill Smith, he would not have told Smith that Jack will attempt to kill him. But then it follows that, by having the power to bring about the nonobtaining of that condition, Jack would have it within his power at t_4 to bring about the nonoccurrence of past events, such as:

Wt_3: God's warning Smith at t_3

Xt_3: Smith's hearing at t_3 a voice telling him that Jack will attempt to kill him

Yt_3: Smith's coming to believe at t_3 that Jack will attempt to kill him, etc.[27]

This, of course, would be a violation of the principle of the fixity of the past in the sense of PFP1. The intuitive principle underlying the last inference is this:

(i) If p, and q is a causally necessary condition (in the circumstances) for p, and $P^*_{X,t}$ (not-q), then $P^*_{X,t}$ (not-p).[28]

There seem to be two ways in which an Ockhamist may try to respond to this objection.

(a) He may diagnose the moral to be drawn from the objection differently. Instead of viewing the objection as showing that power over God's foreknowledge of human free actions may give an agent power over hard facts about the past, he may take it as showing that the following assumption implicit in the objection is false:

(PV) God has it within his power to intervene in the course of events on the basis of his knowledge of future-contingent events.

This inability to intervene, he might say, is the price that an essentially omniscient God must pay if he wants to allow for human freedom.

(b) He may argue that, although God may sometimes intervene in the course of history, he never does so *because* of his knowledge of future-contingent events, but for some other reason that does not involve such knowledge.

[27] Here I am assuming that the occurrence of Wt_3 is causally necessary (in the circumstances) for Xt_3, and that of Xt_3 is causally necessary (in the circumstances) for Yt_3.

[28] In my "A Problem for the Eternity Solution," *International Journal for Philosophy of Religion* (forthcoming), I employ a similar argument against the attempt to solve the conflict between divine foreknowledge and human freedom by conceiving of God as existing outside time.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Both these responses do not seem to me adequate. The first one is theologically implausible. God as portrayed in the Judeo-Christian tradition often intervenes in human history acting in the light of his knowledge of future-contingent events. For example, God appears to Abraham and reveals to him that his descendants will be enslaved and mistreated for four hundred years (Genesis 15:13). He tells Moses that Pharaoh will at first refuse to let the People of Israel go (Exodus 3:19). Knowing that Pharaoh will go in the morning to bathe in the Nile, God orders Moses to meet Pharaoh at the river and to tell him about the first plague (Exodus 7:15). He reveals to the prophet Ahijah that Jeroboam's wife will come to him to consult him about the health of her son, and that she will be coming in disguise (1 Kings 14:5), etc. Thus, the first response seems inconsistent with the traditional account of God's providential activity. It is, of course, open for someone to argue that, on all the occasions on which God allegedly intervened in human history in the light of his knowledge of future-contingent events, those events were inevitable. But note that the objector can make this response only at the expense of violating the notion of divine justice, hence, only at the expense of further attenuating the Judeo-Christian concept of God. For in some of the cases described above (e.g., the one in Genesis 15:13 and that in Exodus 3:19), the agents involved were explicitly punished by God for their deed. Hence, this way out is inadequate.

The weakness of the second response consists in the fact that it renders God's knowledge of future-contingent events completely epiphenomenal. One wonders what role, if any, it plays in God's overall plan for the world. Secondly, if God's knowledge of the future does not play any causal role in his interventions in human affairs, there is the danger that we may end up with situations in which God may tell Smith that Jack will attempt to kill him, even though He knows that Jack does not intend to do so, in which case God would be telling a lie. Needless to say, it is not difficult to conceive of situations that are even more morally repugnant than the foregoing.

It is important to see that the objection from future-contingent causal chains fails when applied to standard-type soft facts, such as (6). If one were to state the objection by replacing 'God knows' with 'Peter correctly believes', it would break down at the point where it assumes that the fact that Peter correctly believed that Jack will pull the trigger at t_5 is a causally necessary condition for the event of his warning Smith. What is in those alternative circumstances a causally necessary condition for the occurrence of that event is not the fact that

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(B1) Peter *correctly* believed at t_1 that Rt_5.
(Rt_5 abbreviates 'Jack pulls the trigger at t_5'.)

but rather the fact that

(B) Peter *believed* at t_1 that Rt_5.

And that is a fact about the past, over which contrary to a fact like

(K1) God foreknew at t_1 that Rt_5.

Jack has no power at t_4. [Unlike (K1), (B) is not a fact concerning which one can say that Jack has it within his power to bring about its nonobtaining, if he has the power to refrain from pulling the trigger at t_5.] Now, someone who adopts a belief model of divine foreknowledge might argue that, when applying the objection to God's case, it is also *not* the fact (K1) that forms a causally necessary condition for God's warning Smith, but rather the fact that

(G1) God believed at t_1 that Rt_5.

But this claim would not be correct, since (K1) is equivalent to (G1). The above remarks suggest a further important difference between God's beliefs and regular soft facts, which is that the causal power of the latter is somehow limited by time. Thus, soft facts such as (4)–(6) cannot, on pain of backward causation, be deemed causally efficacious prior to t_3. They become causally efficacious only after t_3, only after becoming hard facts. Facts like (G) or (G1), on the other hand, are not so limited. Like regular hard facts about t_1, they can be causally efficacious even before t_3.

V

In a recent article, Fischer has also argued that it is a mistake to assimilate a fact like

(G) God believed at t_1 that Qt_3.

to standard-type soft facts, such as

(6) Smith correctly believed at t_1 that Qt_3.

when it comes to the question of their fixity.[29] Fischer adopts the account of the soft/hard fact distinction provided by Hoffman and Rosenkrantz, according to which (G) is classified as a soft fact about t_1. But he maintains that (G) is a very special sort of soft fact about t_1, what he calls a "hard-type soft fact." By this he understands a soft fact whose constitutive property, that of believing that Qt_3, is a temporally genuine property or what he calls a "hard property," relative to t_1. Moreover, he assumes that no human has it within his power at

[29] Fischer, "Hard-Type Soft Facts," pp. 595–9. I discuss this claim of Fischer's more briefly also in "Two Forms of Fatalism," pp. 108–110.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

a time later than t so to act that what is a bearer of a hard property relative to t would not have possessed that property at t. Consequently, hard-type soft facts about t are in his view fixed at times later than t. Fischer offers the following account of the notion of a hard property: a hard property, relative to a given time t, is one which is not soft, relative to t, where by a 'soft property' he understands a property P such that, if *anything* were to have that property at t, it would necessarily follow that some immediate fact obtains after t, in any plausible sense of 'immediate'. So, for instance, temporally relational properties, such as correctly believing that Qt_3, waking up four hours prior to eating lunch, uttering the true sentence 'Jack will sign the contract on January 3', etc., would be examples of soft properties, relative to the pertinent times; whereas those like believing that Qt_3, waking up, eating lunch, etc., would count as hard properties. The motivation underlying this distinction between the two sorts of property is to stress the resemblence between (G) and regular hard facts about t_1 such as that Jack wakes up at t_1, which also have as their constitutive property a hard property. While being sympathetic to this strategy by Fischer of characterizing the asymmetry between (G) and (6), I think that his account of the notion of a temporally genuine or hard property is problematic, and is, therefore, incomplete. To see this, let us assume that

(14) Smith correctly believes at t_1 that it is not the case that Jack will sign the contract at t_4.

Suppose further that Jack has it within his power at t_2 to sign the contract at t_4, but for some reason decides not to do so. Intuitively, (14) is a soft fact about the past (relative to t_2) of the rather soft sort. If Jack had decided to sign the contract at t_4, he could have brought about the nonobtaining of (14). But now let us examine the constitutive property of (14), the property of correctly believing that it is not the case that Jack signs the contract at t_4. On Fischer's account, this property is a hard property, relative to t_1. Clearly, it is not the case that, if anything were to have that property at t_1, it would necessarily follow that some immediate fact obtains at a time later than t_1. For example, we can conceive of a possible world in which the said property is exemplified by Smith at t_1, but in which there are no times later than t_1. In such a world, (14) would be true, although no immediate state of affairs would obtain in it at times later than t_1. Consequently, Fischer would have to treat (14) as a hard-type soft fact about t_1, which is counterintuitive. A further problem for Fischer's account is posed by properties, such as *being true* (as applied to sentences), correctly believing that all ravens are white, or Jack will sign the contract at t_3, etc. These properties, too, would

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

have to be treated by Fischer as hard properties, relative to t_1. (E.g., if Jack raises his arm at t_0, then 'Jack raised his arm at t_0' has the property of *being true* at t_1. But this does not imply that some immediate fact obtains at some time after t_1.) But their exemplification by an object at t_1, may easily yield soft facts of the standard (soft) type, i.e.,

(15) The sentence 'Jack will sign the contract at t_3' is true at t_1.
(16) Smith correctly believes at t_1 that all ravens are white or Jack will sign the contract at t_3.

A further indication that Fischer's account of a hard property is incomplete is provided by the property

P4: being immersed in water and being soluble in water.

This property, which consists of two hard properties—the property of being immersed in water, and the dispositional property of being soluble in water—seems intuitively a hard property, relative to the pertinent time. But on Fischer's account, it would count as a soft property.[30] Keeping in mind that by 'hard property' Fischer understands a temporally genuine property, and assuming that he wants his account of the notion of a temporally genuine property to be more than merely a *stipulative* definition, Fischer owes us an explanation of how is it that by combining two temporally genuine properties we get a temporally nongenuine property. Note that, being a conjunction of two temporally genuine (hard) properties, P4 differs crucially from all of Fischer paradigm examples of soft properties, each of which is either a simple soft property itself, or contains a soft property as a part. Also, unlike Fischer's examples of soft properties, P4, when exemplified by an object at a time t, yields a fact about t that can be analyzed as consisting of two *hard* facts about t.

VI. CONCLUSION

In this paper, I have considered the Ockhamistic response to the problem of divine foreknowledge and human freedom. Having examined various of ways in which this response might be construed, the conclusion I have reached is that none of them seems satisfactory. This conclusion may be viewed as lending support to Pike's main claim that, given the truth of PFP, the existence of an everlasting, essentially omniscient God cannot be reconciled with human freedom.

DAVID WIDERKER
Bar-Ilan University

[30] Another example of such a property would be the property of P5: infallibly believing that Qt_3, which may be viewed as consisting of the property of believing that Qt_3, and the property of being infallible, each a hard property, relative to t_1.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

COMMENTS AND CRITICISM

PARETO UNANIMITY AND CONSENSUS*

T. SEIDENFELD, J. B. Kadane, and M. Schervish[1] have published an important paper on consensual decision making. These authors contend that the consensual preference among available options ought to preserve those comparisons among available options which all parties to consensus agree in making. Pareto unanimity ought, in this sense, to be a constraint on consensual decision making.[2]

Seidenfeld, Kadane, and Schervish then consider situations where there are two agents (they speak of Dick and Jane) concerned to fix on a consensus, who face so called "horse lotteries" in the technical sense of F. J. Anscombe and R. Aumann[3] over two states. If Dick and

* Thanks are due to Teddy Seidenfeld for important error detection.

[1] "On the Shared Preferences of Two Bayesian Decision Makers," this JOURNAL, LXXXVI, 5 (May 1989): 225–244.

[2] Consensual decision making is a species of group decision making where a choice among a set of options is based on the shared agreements among the parties to consensus concerning (i) what is taken for granted as settled or certain, (ii) probability judgments, and (iii) values or utilities. Consensual decision making ought to satisfy minimal conditions of rationality imposed on individual decision makers. This requirement need not hold for other forms of group choice where each member of the group makes a decision according to his respective beliefs and desires and the "group" choice is the product of these individual choices.

Seidenfeld, Kadane, and Schervish agree with the view I favor that, if the parties to consensus order the available options according to different preference rankings over all horse lotteries generated by a given set of states and prizes (see footnote 3), not only should the consensus respect their shared agreements, it should not impose a preference over the horse lotteries where the parties to consensus disagree. The consensus should be represented by all weak orderings that preserve the shared agreements. See footnote 4.

[3] "A Definition of Subjective Probability," *Annals of Mathematical Statistics*, XLIX (1963): 199–205. A roulette lottery over a set of prizes is a stochastic process which when implemented yields prizes from the given set with specific objective or statistical probabilities or chances. A horse lottery relative to a specified list of states and prizes is an option which, for each state, yields a roulette lottery on the prizes as the payoff. Each horse lottery can be represented as a function from the states to the roulette lotteries. Given a weak preference ordering over all horse lotteries relative to a finite set of states and a finite set of prizes satisfying von Neumann-Morgenstern conditions on roulette lotteries over prizes and also over horse lotteries, a dominance or monotonicity axiom over horse lotteries and an axiom that renders each mixture of horse lotteries equal in value to a corresponding pure horse lottery, there is a unique probability p over the states and a state independent utility u unique up to a positive affine transformation such that the expected utilities of the horse lotteries according to (p, u) represent the weak ordering. In effect, the axioms ensure that the preference ordering satisfies the requirements of strict Bayes rationality when the states are probabilistically independent. If in addition to what is

0022-362X/90/8709/481–492
© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

 Digitized by Arya Samaj Foundation Chennai and eGangotri

Jane share the same probability distribution over the states but different utility assignments to the prizes, Pareto unanimity requires that every utility function defined over the prizes which is a permissible resolution of the differences between Dick's utility and Jane's must be a positive affine transformation of a weighted average of Dick's utility and Jane's utility. The set of permissible utility functions will be the set of all such weighted averages. It satisfies the convexity requirement on permissible utility functions which I advocated in "On Indeterminate Probabilities."[4] If Dick and Jane share the utility but have different probabilities, Pareto unanimity requires that every permissible probability be a weighted average of Dick's and Jane's. Again, the set of such permissible probabilities satisfies the convexity requirement.

Consider, however, the case where both the probabilities and the utilities of Dick and Jane differ in some respect. Seidenfeld, Kadane, and Schervish prove the remarkable result that the only two probability-utility pairs that preserve Pareto unanimity are those representing the views of Dick and of Jane. Hence, these are the only two permissible probability-utility pairs if we insist that Pareto unanimity is obeyed in consensus.

This result conflicts with the proposals I have made for such cases, according to which one should consider as permissible all weighted averages of Dick's probabilities and Jane's and all weighted averages of Dick's utilities and Jane's, and consider all methods of determining expected utilities to be permissible which are based on a probability and a utility permissible from these two sets.

There is no gainsaying the results demonstrated by Seidenfeld,

---

implied by the axioms, we assume that utilities are state-independent, strict Bayes rationality requires that expected utility be maximized according to the probability-utility pair given in the representation theorem.

[4] This JOURNAL, LXXI, 13 (July 18, 1974): 391–418. See also *The Enterprise of Knowledge* (Cambridge: MIT, 1980), and *Hard Choices* (New York: Cambridge, 1986). In the context of evaluating horse lotteries (on the assumption of state-independent utilities), a weak ordering of the horse lotteries is a permissible preference ranking of the horse lotteries if and only if there is a permissible probability-utility pair representing that ordering. A probability distribution over the states is permissible if and only if it is an element of some permissible probability-utility pair. The same for the permissibility of a utility function over the prizes. Given a subset of the horse lotteries as a set of feasible options, an element of the set is E-admissible if and only if it maximizes expected utility in the feasible set according to some permissible probability-utility pair. In my proposals, I require (a) that if p and u are permissible than so is the pair (p, u), (b) that the set of permissible probabilities is convex, and (c) the set of permissible utilities is convex. Seidenfeld, Kadane, and Schervish contend that these requirements are acceptable for consensual decision making only in certain restricted contexts.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Kadane, and Schervish. What may be questioned is whether unqualified Pareto unanimity is required in consensual decision making as these authors seem to think it is. The attractions of Pareto unanimity are undeniable. But there is a problem. To appreciate the nature of the problem, a numerical illustration (a minor variation of theirs) of their result may be useful.

Suppose Dick and Jane jointly face a choice between two "horse lotteries" in the technical sense of Anscombe and Aumann. Let c^*, c, and c_* be three prizes. Dick and Jane agree that c^* is better than c, which is better than c_*. Let rl be a roulette lottery that yields c with a chance of 19/30, c^* with a chance of 8/300, and c_* with a chance of 102/300. The two options faced by Dick and Jane are given in table 1.

Table 1

	s	$\sim s$
Option A	rl	rl
Option B	c_*	c

s and $\sim s$ are two "states" about whose truth Dick and Jane are uncertain. Dick assigns probability 0.1 to s (and 0.9 to $\sim s$) and Jane assigns probability 0.3 to s (and 0.7 to $\sim s$). Letting the utility of c^* be 1 for both agents and for c_* be 0 (as we may do without making any substantial assumption), Dick's utility for c is 0.1 and Nancy's is 0.4. Computation reveals that the expected utilities assigned by Dick to option A and option B are both equal to 0.09. Nancy's expected utilities for these two options are both 0.28.

Thus, in spite of the differences in their probability and utility judgments, Dick and Nancy agree that A and B are to be equivalued.

Because of this agreement, Seidenfeld, Kadane, and Schervish conclude that the consensual point of view ought to preserve this equivaluation as well, just as Pareto unanimity requires. That is to say, any permissible comparison of A and B should rank them together.

We are assuming, however, that consensus is rational in the sense that every permissible ranking of the options should be representable by a pair of probability assignments to the states and utilities to the prizes such that the evaluation of the options is representable in terms of expected utilities.

Combining this requirement of Bayes rationality with Pareto unanimity implies that the expected-utility representations should rank

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

A and B together. This implies that any permissible pair (p, u) of probability and utility functions should satisfy the following condition:

(i) $(19/30)u(c) + 8/300 = p(\sim s)u(c)$

Let rl_1 be a roulette lottery yielding c^* with a chance 0.1 and c_* with a chance (0.9). Let the payoffs for options C and D be as given in table 2.

Table 2

	s	$\sim s$
Option C	rl_1	rl_1
Option D	c^*	c_*

Ron and Nancy agree that C is not better than D. Pareto unanimity and Bayes rationality then require that the permissible probability-utility pairs should restrict probabilities to the interval between 0.1 and 0.3.

Let rl_2 yield c^* with chance 0.4 and c_* otherwise. Consider then the three options consisting of option C, option E which yields c for sure, and option F which yields rl_2 for sure. Ron and Nancy agree that C is not better than E, which in turn is not better than option F. To preserve these shared agreements, the utility of c must be restricted to the interval from 0.1 to 0.4.

If we seek to preserve the shared agreements in preferences between C, D, E, and F, we now see that any probability-utility pair (p, u) which assigns probability to s in the range from 0.1 to 0.3 and assigns utility to c in the range between 0.1 and 0.4 will satisfy our requirements. If we seek to preserve the shared agreement concerning comparisons between A and B as well, however, the only probability-utility pairs from that set which satisfy our requirements are the pairs (0.1, 0.1) and (0.3, 0.4) representing Dick's and Jane's points of view, respectively.

An implication of this is that, given any pair of horse lotteries H and H' on the pair of states s and $\sim s$, such that Dick prefers H to H' and Jane prefers H' to H, there is no permissible evaluation of these options which ranks them together. There are no potential compromises.

Seidenfeld, Kadane, and Schervish generalize the result for this particular example and show that, when two Bayes-rational agents

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

differ in their probabilities and utilities over a set of horse lotteries, the set of potential resolutions that preserve shared agreements as Pareto unanimity does contains no compromises.

Seidenfeld, Kadane, and Schervish exhibit no doubt that unrestricted Pareto unanimity is a *conditio sine qua non* of consensus as shared agreement. They are prepared to abandon the inclusion of compromises among the potential resolutions. This means that, even when choosing between roulette lotteries (i.e., opportunities where chances are given), Dick and Jane must not countenance any compromise potential resolutions as long as they differ in both their probability and utility judgments and no matter how slight these differences might be. Consider, for example, a choice between receiving prize c for sure and receiving a roulette lottery that yields c^* with a chance of 0.25 and c_* with a chance 0.75. Dick prefers the roulette lottery to c but Jane prefers c to the roulette lottery. According to the position taken by Seidenfeld, Kadane, and Schervish, if Dick and Jane also have the option of flipping a fair coin to decide the issue, this option ought not to be counted as E-admissible. According to Dick's view, it is inferior to the roulette lottery. According to Jane's, it is inferior to the prize c.

Prohibiting such mixed options as inadmissible is surely an extreme view to take. I am not contending that adopting the mixed option is rationally obligatory. But it is not rationally prohibited either. Such prohibition is, however, an implication of the Seidenfeld, Kadane, and Schervish position if E-admissibility is necessary for admissibility.

To be sure, the Seidenfeld-Kadane-Schervish approach removes the prohibition if there is complete unanimity on either the probability or on the utility function. It is a remarkable and, perhaps, objectionable aspect of their approach that it allows and, indeed, mandates the permissibility of compromises if there is complete agreement on either probability or on utility, but prohibits compromises otherwise.

These reservations would, perhaps, be worth little were unrestricted Pareto unanimity a compelling requirement on consensus as shared agreement. And at first glance, it does seem obviously mandatory. What is a consensus as shared agreement if it does not preserve shared agreements?

A second glance, however, reveals that the case for Pareto unanimity is not only not compelling but is clearly untenable—at least for anyone who thinks that, in the evaluation of horse lotteries, "Bayesian" rankings representable by probability-utility pairs are the only potential resolutions in consensus. Since there is no disagreement

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

between Seidenfeld, Kadane, and Schervish and myself on this score, we may leave this Bayesian assumption unquestioned.

Dick and Jane differ in their credal probabilities and utilities. Seidenfeld, Kadane, and Schervish do not tell us whether their differences are accompanied by a difference in their corpora (i.e., standards for serious possibility[5]). I myself do not think that the status of Bayes rationality and Pareto unanimity ought to depend on whether the participants disagree with respect to the corpus or not. In particular, either Pareto unanimity ought to apply in both cases or fail to apply in both cases. Whether Seidenfeld, Kadane, and Schervish agree is unclear. The fact that they overlook the case where participants in a joint decision differ with respect to corpus as well as with respect to probability and utility judgment suggests, however, that they may agree that the general conditions on consensus ought to cover both cases.

Consider then the case of Ron and Nancy, who resemble Dick and Jane in their credal probabilities and utilities as envisaged in our numerical example. Unlike Dick and Jane, it is clear that they also disagree with respect to the corpora to which they are committed as standards for serious possibility. Ron's corpus K_1 is the set of logical consequences of e and Nancy's corpus K_2 is the set of logical consequences of $\sim e$. Let us suppose, moreover, that Ron and Nancy agree in their confirmational commitments. That is to say, Ron agrees with

[5] For discussion of serious possibility, see *Enterprise of Knowledge*, chs. 1–3, and *Decisions and Revisions* (New York: Cambridge, 1984, chs. 8, 9, 11, 12, and 14).

Characterizing X's corpus of knowledge as X's standard for serious possibility, as I have been accustomed to do, seems to be subject to a certain misunderstanding. Let us say that X fully believes that h when there is no serious possibility according to X that h is false. Then, according to what I have said about knowledge, everything X knows, X fully believes. What about the converse? From X's point of view, everything X fully believes is true so that there is no difference between what X fully believes and what X knows. From Y's point of view, however, there can be a substantial difference between what X knows and what X fully believes. Many critics have refused to appreciate my emphasis on this point.

Although I have characterized knowledge in terms of its function in inquiry, I have never offered this characterization as a necessary and sufficient condition for knowledge—except from the point of view of the knower. Is there a specification of necessary and sufficient conditions for knowledge which is the same for all agents. Yes! X knows that h if and only if X fully believes that h and it is true that h. A body of knowledge is, therefore, an error-free standard for serious possibility. From X's point of view, X's standard for serious possibility is error-free. The distinction between knowledge and full belief collapses as before.

What is ruled out is any appeal to a justificatory or explanatory condition on knowledge. Questions of justification arise when changes of point of view are being contemplated. With Charles Peirce, I say that, when there is no "living doubt," inquiry (and, hence, justification) is not demanded.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Nancy that, relative to K_2, 0.3 is the appropriate credal probability assignment to s. And Nancy agrees with Ron that, relative to K_1, 0.1 is the appropriate credal probability assignment. Finally, Ron and Nancy agree that, relative to the corpus K according to which both e and $\sim e$ are serious possibilities, the credal probability of e should be x and the credal probability of $\sim e$ should be $1 - x$.

Under any familiar understanding of consensus as shared agreement, the parties to consensus should move to a corpus which preserves just those judgments of serious possibility and impossibility which they share in common, and which counts every proposition a serious possibility when parties to the consensus differ as to its status. Without agreement on the corpus, there is no consensus as shared agreement.

But if Ron and Nancy move to K and assign hypothesis e credal probability x, hypothesis s receives credal probability $0.1x + 0.3(1 - x)$. If we assume, as we consistently can, that neither Ron nor Nancy changes their utilities for roulette lotteries relative to K from what they were initially, then the probability-utility pair that represents consensus must rank options A and B in violation of unrestricted Pareto unanimity. Ron and Nancy both equivalue A and B. They both weakly prefer F over E, weakly prefer E over C, and weakly prefer D over C. As we have seen, the only probability-utility pairs that satisfy these consensual valuations are Ron's and Nancy's. On the other hand, as long as x is distinct from 0 and 1, the probability-utility pair representing consensus relative to the consensus corpus is distinct from either of these pairs. Options C, D, E, and F will continue to be ranked in the way required by Pareto unanimity relative to the consensus corpus, but A and B will not.

Observe that this objection to the Pareto-unanimity condition cannot be dismissed so readily as the implication that there are no compromises by saying that Pareto unanimity is definitory of the conception of consensus as shared agreement. Insisting that compromises of conflict be permissible does carry considerable plausibility. Still one may well doubt that, whatever the attractions of compromise may be, it is ingredient in our understanding of consensus as shared agreement. But the requirement that consensus as shared agreement presupposes that potential resolutions of the conflicting views relative to the consensus corpus should be the permissible potential resolutions of the conflicting views of the parties to consensus relative to the corpora they actually adopt seems noncontroversial as a requirement on consensus as shared agreement. It requires that, in such consensus, the permissible probability-utility

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

pairs should be defined over the space of possibilities determined by the consensus corpus. That should go without saying.[6]

Strictly speaking, I have begged the question against Seidenfeld, Kadane, and Schervish. But at this point, I must plead with the reader to consider whether a point of view is a consensus as shared agreement unless there is a single corpus relative to that consensus —the consensus corpus of shared agreements as to what is certain and unless the consensual evaluation of horse lotteries is relative to the consensus corpus.

To sustain my own affirmative answer to this question, however, it is desirable to appeal to more than presystematic conceptions of consensus as shared agreement, as I have done. After all, the Pareto-unanimity principle also seems presystematically compelling. In order to support my rejection of Pareto unanimity, it is desirable that a more modest version of the principle be suggested which can plausibly be said to accommodate what is presystematically attractive about Pareto unanimity.

Ron's corpus is K_1. Nancy's K_2. Ron can recognize that, were he to endorse Nancy's probability judgments relative to her corpus while keeping his utility function for roulette lotteries in tact, his ranking of the elements of horse lotteries would differ from what it currently is. In point of fact, it would be Nancy's under the conditions we have envisaged. Some of the comparisons of horse lotteries would remain the same, however, relative to (0.3, 0.1) as they are relative to (0.1, 0.1). These comparisons of value we shall call Ron's robust preferences for the purpose of consensus with Nancy. Nancy like Ron has robust preferences for the purpose of consensus with Ron. With the notion of robust preferences for the purpose of consensus, we can then identify those robust preferences which Ron and Nancy share. We may speak of robust consensual strict preference, indifference and weak preference in a sense paralleling the corresponding no-

[6] In the case envisaged, relative to the consensus corpus neither the probability-utility pair (0.1, 0.1) or the pair (0.3, 0.4) is a potential resolution of the conflict. This does not imply that the differing points of view of Ron and Nancy are ignored. These two probability-utility pairs represent the viewpoints of Ron and Nancy relative to distinct corpora. We require only that their views relative to the consensus corpus be considered.

If these two probability-utility pairs were permissible, it would then turn out that, according to the consensus, the assignment of probability 1 to e and of probability 0 to e would both be permissible. This cannot be right relative to the consensus corpus. We have assumed that Dick and Jane agree that the credal probability for e is x relative to the consensus corpus. This might be 0 or it might be 1 (although neither is very plausible). But it cannot be both.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tions of consensual preference. Finally, we obtain the idea of a robust consensual quasi ordering.

Observe that the robust consensual quasi ordering is induced by the probability-utility pairs for Ron and Nancy relative to their current points of view whether they share the same corpus or not. We may distinguish that robust consensual quasi ordering from the robust consensual quasi ordering that they would endorse were they to move to the consensus corpus. We shall call the latter the *epistemically robust consensual quasi ordering.*

With these ideas in place, the Pareto-unanimity condition may be modified as follows:

> *Robust Pareto Unanimity:* Every potential resolution is a consistent extension of the robust consensual quasi ordering over the set of horse lotteries.

Robust Pareto unanimity and Bayes rationality may now be taken to be necessary conditions on potential resolutions relative to the probability-utility pairs for the parties to the dispute, given the corpora they actually adopt. These two conditions are jointly sufficient.

If we assume that the probability judgments endorsed by Ron and Nancy relative to their current corpora are derivable from the consensus corpus by conditionalization and Bayes theorem upon adding the information present in Ron's (Nancy's) current corpus but not in the consensus corpus, it is clear that all robust comparisons of horse lotteries relative to the consensus corpus are robust relative to the current corpora, though the converse does not hold except when Ron and Nancy share the same corpus.

Consequently, since Ron and Nancy do not share the same corpus, our current proposal does not imply that Ron's and Nancy's probability-utility pairs are permissible.

On the other hand, if Dick and Jane share the same corpus, our conditions do imply that both of their probability-utility pairs are permissible. Hence, these evaluations of horse lotteries are potential resolutions in this case. We get mutual respect for the views of the participants in consensual choice in the sense that these views count as permissible precisely when we want it—to wit, when their current corpora are one and the same and, hence, qualify as the consensus corpus (or where the probability-utility pairs relative to the consensus corpus are identical with those relative to the current corpora in all relevant respects so that disagreements in corpus are irrelevant).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

What about the permissibility of compromises? Whether or not Ron and Nancy share a corpus in common, Bayes rationality and restricted Pareto unanimity imply that all weighted averages of the expected-utility functions defined by their probability-utility pairs represent potential resolutions. To see this, consider that C is robustly strictly preferred to D relative to Ron's and Nancy's views. Because of this, this preference is preserved by the pairs (0.1, 0.4) and (0.3, 0.1). Hence, they are also satisfied by all probability-utility pairs (p, u) where p is a weighted average of 0.1 and 0.3 and u is a weighted average of 0.1 and 0.4.

If a is strictly preferred to a' according to Ron and a' is strictly preferred to a according to Nancy, then either a and a' are equipreferred according to (0.1, 0.4) or according to (0.3, 0.1) or according to neither. If one of the first two alternatives obtained, the compromise condition is satisfied, if the third alternative is satisfied, then one of these pairs ranks a over a' or a' over a. Either way, there is a conflict between the ranking induced by this pair and either Ron's or Nancy's ranking. But now the conflict obtains between pairs that contain either a probability or a utility in common and, as we saw earlier on, the compromise condition is satisfied by the unrestricted Pareto unanimity and a fortiori the robust Pareto-unanimity condition in that case.

This does not mean that a compromise potential resolution will be permissible in consensus. For this to obtain, either Ron and Nancy must share the same corpus or, if they have different potential corpora, the set of potential resolutions of their conflicts relative to the consensus corpus must include compromise potential resolutions relative to the actual corpora for each paired comparison between horse lotteries. This need not, in general, obtain.

Thus, it has been shown that robust Pareto unanimity is compatible with the requirement that there be permissible compromises of conflict even though the unqualified Pareto-unanimity condition is not. It has also been shown that robust Pareto unanimity is compatible with the requirement that, in consensus as shared agreement, judgments should be made relative to the consensus corpus. Unqualified Pareto unanimity is not compatible with this requirement.

Seidenfeld, Kadane, and Schervish, in effect, point out that weakening Pareto unanimity to robust Pareto unanimity fails to make a distinction between seeking consensus as shared agreement when Dick and Jane are the participants, when Tom and Mary are the participants and where Tom's evaluation of the given options is representable by (0.1, 0.4) and Mary's by (0.3, 0.1), and when Dick, Jane,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Tom, and Mary all join it. In these three cases, the pairs of acts about whose ranking all parties to consensus agree are different. But the differences occur only with respect to those agreements which lack robustness. With respect to robust agreements, all three cases are the same and, hence, the set of potential resolutions on my view are the same. Seidenfeld, Kadane, and Schervish seem to think that this is a defect in my view. It fails to make fine-grained enough distinctions. They see vice in neglecting nonrobust consensual preferences in consensus. Where they see vice, I see virtue.

To see the virtue, suppose that Ron, Nancy, George, and Barbara agree that 0.1 is appropriate to use as the probability of s relative to e and 0.3 is appropriate to use relative to $\sim e$. Ron and George agree that e is true and Nancy and Barbara that $\sim e$ is true. Now, just as Ron and Nancy should reach consensus as shared agreement by moving to suspense between e and $\sim e$, so should George and Barbara and so should all four do when they are seeking to make a joint consensual decision. If they all agree that x is the probability of e in the state of suspense, the consensus probability-utility pairs will be the same in all three cases. If they recognize all values of x in some range as permissible, the same will be true. In short, there will be no difference between the three cases of consensus when the parties do not share a common corpus to begin with. If the status of Pareto unanimity is the same whether participants to a consensus share the same corpus or not, as I am maintaining, there should be no difference between the cases of consensus when the parties do share a common corpus at the outset.

I concede this much to the unqualified devotion to Pareto unanimity advocated by Seidenfeld, Kadane, and Schervish. A Catholic and a Communist might face a choice between two policies and agree that one is better than the other even though they differ in probability and utility. They might agree to act in concert without seeking a consensus as shared agreement. They might rest content with a political consensus. But such a political consensus is not consensus as shared agreement. Indeed, the Catholic and Communist both view the joint action as the action of two agents each acting separately to promote his or her personal agenda as in a game. Both prefer a national health-insurance plan with a prohibition of aid for abortions to no health plan at all. Each will vote for the plan. But their both voting is not a group action grounded on consensus as shared agreement. The alleged group action is the implementation of the national-health plan which by hypothesis is the outcome of their both voting for it. In our example, it turns out that a form of pareto

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

unanimity will be observed. The Catholic and the Communist both prefer the health plan being implemented than not. And their "votes" will result in its implementation. But as is well-known, there are game-like situations (e.g., prisoners' dilemmas) where the "votes" or choices of the participants will yield Pareto-inferior outcomes even though the players are fully rational. In any case, when group action is the outcome of the independent decisions of members of the group, there is no failure of group rationality. There is no group agent to fail.

Seidenfeld, Kadane, and Schervish start with the premise that the kind of group decision making they are considering is not of the game-like sort just mentioned but belongs to the category of social-choice problems where the decisions made should meet canons of rationality applicable to individual choice. Yet, the result they come up with according to which the only resolutions of the conflict between Dick and Jane are the views of Dick and Jane is, for reasons I have already indicated, unacceptable as an account of consensual decision making. It remains an open question whether there are any contexts of group behavior where their requirements may plausibly be imposed.

Seidenfeld, Kadane, and Schervish have produced an impossibility theorem which is at once a surprising and novel contribution to the Arrovian tradition and a fertile source for philosophical reflection. We should be grateful to them for having opened up new avenues for exploration in theories of individual and group decision making. Nonetheless, I think their unrestricted endorsement of Pareto unanimity and abandonment of the requirement that in consensus compromises should be countenanced as potential resolutions is misguided. To my way of thinking, the important lesson to be drawn from their theorem is that adherence to Pareto unanimity is not all it is cracked up to be.

ISAAC LEVI

Columbia University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY
VOLUME LXXXVII, NO. 10, OCTOBER 1990

"BUT WOULD THAT STILL BE ME?" NOTES ON GENDER, "RACE," ETHNICITY, AS SOURCES OF "IDENTITY"*

IF you had asked most Anglo-American philosophers twenty-five years ago what conditions someone had to meet in order to be (identical with) me, they would, no doubt, have taken this (correctly) to be a conceptual question, and (incorrectly) inferred that it was to be answered a priori by reflection on the properties whose presence would have led them to *say* that an imagined entity was Anthony Appiah. Since there are hardly any properties of persons whose absence we cannot intelligibly imagine, it was tempting to conclude that there was something odd about the very question.

In these enlightened post-Kripkean times, we think we know that it was the way of trying to answer the question which was odd. For we now think that the question whether (as we are likely to put it) some individual in a possible world is AA is an a posteriori question about a real essence. Some believe not only that this is a question about real essences, but that we know its answer: that the real essence of a person is the chromosomal structure produced by the coition of his actual parents, a thesis that is the biological fleshing out of the metaphysical doctrine of the necessity of origins.

These are important issues in the semantics, metaphysics, and logic of identity, and they are centrally concerned with the identification of individuals across (metaphysically) possible worlds. But it seems to me that there is an equally important set of questions that recent theorizing has left to one side, a set of questions that can also be raised by asking, about a possible individual, "But would that still be

*To be presented in an APA symposium on Gender, Race, Ethnicity: Sources of Identity, December 30. Maria C. Lugones will be co-symposiast, and Thomas E. Wartenberg will comment; see this JOURNAL, this issue, 500–7 and 508–9, respectively, for their contributions.

0022-362X/90/8710/493–9

© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

me?" I want to argue that there is a sense of this question which is best answered in the "old-fashioned" conceptual way; and to get at what I have in mind, nothing could provide a better starting point than questions about "race," ethnicity, gender, and sex.

Consider, for the purposes of an initial example, the possibility that I might have been born a girl. Someone convinced of the chromosomal account of individual identity, and convinced, too, that what it is to be biologically female or male is to have the appropriate chromosomal structure, will argue that this is only an apparent possibility.[1] A female person could have been born to my parents when I was, if a different sperm and egg had met: but she would not have been me. It will be false, in this view, that I could have been born a woman.

I am prepared to concede all this for the purposes of argument; but there is a different question I might want to consider about a different possibility. Might I not, without any genetic modification, have been raised as a girl? This sort of thing certainly can happen; as when, for example, surgeons engaged in male circumcision remove the whole penis in error: rather than face a child with what—in our society—is bound to be the trauma of growing into a man without a penis, surgeons will often, in such circumstances, remove the testes from the abdomen, construct a facsimile of the female external genitalia, and ask the parents to bring the child back for hormone therapy in time to manage a facsimile of female puberty. If the good doctor who circumcised me had made such a mistake, could not I—this very metaphysical individual here—have been raised with a feminine (social) gender even though, on the chromosomal essentialist view, I was still of the male (biological) sex?[2]

My claim in this paper is that, while there may be a sense of the question, "Would that have been me?", under which the answer to this question is 'yes', there is another, intelligible reading under which it could, surely, be 'no'.

To get at that reading, consider the—admittedly, very different—possibility that I might seek to have a sex change, prior to which I could consider our guiding question about the possible future social female this metaphysical individual would then become. "Would that still be me?", I could ask. Now it seems to me that I can give either of

[1] Given the underlying biology, this is too simple a view: I assume it in this paper as a first approximation.

[2] I shall use male and female for biological identities (sex) and feminine and masculine for social identities (gender). See the remarks on the topic of XX and XY genotypes below.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

two answers here, and that which answer I should give depends in large part on how central my being-a-man—my social masculinity and, perhaps, my possession of the biological appurtenances of maleness—is, as we would ordinarily say, to my identity. And it is in exploring this sense of the term 'identity' that we can come to learn why it is that there is a sense to this question—I shall call it the ethical sense—in which I may chose to answer it in the negative.

To say that I may chose is to speak loosely. The issue is not really a matter of choice. What answer I should give to the question understood this way depends on how central my being-a-man is to my identity, not on how central I choose to make it. Transsexuals will surely answer in the affirmative; they often say that they were always of the "other" sex all along. For the chromosomal essentialist, this will be false. But a transsexual might (after reading Saul Kripke) come to conclude that what he or she really had in mind was the different thought that his or her real identity, in the sense of the term I am now trying to explore, was that of the sex into which he or she was not born. And if I were a transsexual convinced of this I would say, contemplating the feminine person that I might become, "Yes, that would be me; in fact it would be the real me, the one I have always really been all along."

But what I am actually inclined to say is: "No. A sex-change operation would make of this (metaphysical) person a different (ethical) person." And so there is a sense in which she would not be me.

As many people think of them, sex (female and male, the biological statuses) and gender (masculine and feminine, the social roles) provide the sharpest models for a distinction between the metaphysical notion of identity that goes with Kripkean theorizing and the notion of identity—the ethical notion—that I am seeking to explore. I say 'as many people think of them' because the real world is full of complications. Not every human being is XX or XY. And there are people who are XY in whom the indifferent gonad was not prompted to form the characteristic male external genitalia; people whom it seems to me odd to regard as "really" biological males. Just as it would be odd to treat an XX person with male external genitalia, produced as the result of a burst of testosterone from a maternal tumor, as "really" biologically female. Once you have an inkling of how messy the real world of the biology of the reproductive organs is, you are likely, if you are wise, to give up the idea that there are just two biological sexes into which all human beings must fall. And this is important because most people do not make the distinctions (or know the facts) necessary to appreciate this, and thus have thoughts

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

about what it is to be a man or a woman which involve concepts that essentially presuppose falsehoods about how people biologically are. Before someone has made a sex-gender conceptual distinction we cannot always say whether what these thoughts were about was one or the other: there are, so to speak, thoughts that no one who *had* made these distinctions could have.

But the general point can be made in cases far from the biological hard cases: if you consider a straightforward case of an XY biological infant, born with standard male internal and external genitalia, who is assigned a feminine gender as the result of early loss of his gonads, it is clear that such a person can agree to a Kripkean "metaphysical" identification as a biological male and insist on the centrality to her of her feminine-gender identity, on being, so to speak, ethically a woman. But before I say more about what this means, it will help to have a couple of rather different cases before us.

Take next, then, so-called "racial" identity. Here the biological situation is much worse than in the case of sex. No coherent system of biological classification of people—no classification, that is, that serves explanatory purposes central to biological theory—corresponds to the folk-theoretical classifications of people into Caucasian, Negro, and such. This is not, of course, to deny that there are differences in morphology among humans: people's skins do differ in color. But these sorts of distinctions are not—as those who believe in races apparently suppose—markers of deeper biologically-based racial essences, correlating closely with most (or even many) important biological (let alone nonbiological) properties. I announce this rather than argue for it, because it is hardly a piece of biological news, being part of a mainstream consensus in human biology. This means that here we cannot make use of an analog of the systematic sex-gender distinction: the underlying biology does not deliver something that we can use, like the sex chromosomes, as a biological essence for the Caucasian or the Negro.

But this does not mean that people cannot have ethical identities tied up with being, say, Euro- or African- or Asian-American; what it does mean, given that such identities often presuppose falsehoods about the underlying biology, is that, once the facts are in, a different theoretical account of those identities is required. From an external point of view, we can construct an account of what it is that people take to be grounds for assigning people to these racial categories. We can note that they are supposed to be asymmetrically based on descent: that "whites" in America are supposed to have no non-"white" ancestry, but that "blacks" and Asians may have

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

non-"black" and non-Asian ancestry. But from the point of view of people whose ethical identity is at stake, it is not going to be enough simply to remark how others classify them. And to see this we can return to our guiding question.

Let us suppose[3] that an American of African descent could be offered the possibility of losing all the morphological markers that are associated in this society with that descent. Her skin is lightened, her hair straightened, her lips thinned: she has, in short, all the services of Michael Jackson's cosmetic surgeon and more. Surely, in contemplating this possibility, she could ask herself whether, once these changes had occurred, the resulting ethical person "would still be me." And, so far as I can see, almost everyone who does contemplate this question in our society is likely to judge that, whether or not these changes are desirable, the answer here must be 'yes'.

I am asserting here, therefore, a contrast between our attitudes to (ethical) gender and (ethical) "race." I suggest that we standardly hold it open to people to believe that the replacement of the characteristic morphology of their sex with a (facsimile) of that of the other (major) one would produce someone other than themselves, a new ethical person; while the replacement of the characteristic morphology of their ethical "race" by that of another would not leave them free to disclaim the new person. "Racial" ethical identities are for us—and that means something like, us in the modern West—apparently less conceptually central to who one is than gender ethical identities.

That this is so does not entail that being-an-African-American cannot be an *important* ethical identity: it is a reflection, rather, of the fact that ethical identity is not a matter of morphology, that skin and hair and so on are simply signs for it. Such an identity is, as we ordinarily understand it, exactly a matter of descent: and nothing you do to change your appearance or behavior can change the past fact that your ancestors were of some particular origin. Nevertheless, even for those for whom being-African-American is an important aspect of their ethical identity, what matters to them is almost always not the unqualified fact of that descent, but rather something that they suppose to go with it: the experience of a life as a member of a group of people who experience themselves as—and are held by others to be—a community in virtue of their mutual recognition—and their recognition by others—as people of a common descent.

[3] As George Schuyler actually did in his engaging moral fantasy, *Black No More* (New York: Negro Universities Press, 1931).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

It is a reasonable question how such "racial" identities differ from those we call "ethnic." What matters about the identity of, say, Irish-Americans—which was conceived of racially in the nineteenth century in North America—is that it, like an African-American identity, involves experiences of community in virtue of a mutual recognition of a common descent. What differentiates Irish-American from African-American identity, as understood in these United States, is that it is largely recognized nowadays that what flows from this common descent is a matter of a shared culture. People of Irish-American descent adopted and raised outside Irish-American culture are still, perhaps, to be thought of as Irish-Americans; but they have a choice about whether this fact, once they are aware of it, should be central to their ethical identities, and their taking it as central would involve them in adopting certain cultural practices. Someone who refuses to do anything with the fact of their Irish-American descent—who fails to acknowledge it in any of their projects—is not generally held to be inauthentic; is not held to be being unfaithful to something about herself to which she ought to respond. So far as I can see, by contrast, African-Americans who respond in this way fall into two categories, depending on whether or not their visible morphology permits them to "pass," permits them, that is, to act in society without their African ancestry's being noticed.

If they cannot pass, they will often be thought of as inauthentic, as refusing to acknowledge something about themselves that they ought to acknowledge, though they will not be thought to be dishonest, since their morphology reveals the fact that is being denied. If they can pass, they will be thought of by many, as being not merely inauthentic but dishonest. And while they may have prudential reasons for concealing the fact of their (partial) African descent, this will be held by many to amount to inauthenticity, especially if they adopt cultural styles associated with "white" people.

Now, so far as I can see, these differences between the identities that we think of as "racial" and those which we think of as "ethnic" cannot be made intelligible without adverting to certain (false) beliefs. Someone who conceals the fact of an African ancestry in his social life quite generally is held to be inauthentic, because there is still around in the culture the idea that being (partially) descended from black people makes you "really" black—in ways that have ethical consequences—while being descended from Irish stock merely correlates roughly with a certain cultural identity. If "races" were biologically real, this would, perhaps, begin to be a possible distinction; though it would require further argument to persuade me that

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ethical consequences flowed from membership in races. But since they are not, this distinction seems, as I say, to require a distinction that someone apprised of the facts should just give up.[4]

That "race" and gender have interestingly different relations to metaphysical identity should not obscure the fact that as ethical identities they have a central importance for us. What this means is, presumably, something like this: that for us, in our society, being-of-a-certain-gender and being-of-a-certain-race are for many people facts that are centrally implicated in the construction of life plans. To ignore one's race and one's gender in thinking about the ethical project of composing a life for oneself requires, in many minds, a kind of ignoring of social reality which amounts to attempting to fool oneself; and that is part of what is involved in the thought that passing for the "wrong" gender or race involves a certain inauthenticity.

We construct ethical identities—woman, man, African-American, "white"—in ways that depend crucially on false beliefs about metaphysical identities; something like each of them could be reconstructed out of other materials. But if we were to live in a society that did not institutionalize those false metaphysical beliefs, it is unclear that the project of reconstruction would be an attractive one. In a truly nonsexist, nonracist society, gender, the ethical identity constructed on the base of sexual differences, would at least be radically differently configured, and might, like "race," entirely wither away; ethnic identities, by contrast—and this is something an African-American identity could become—seem likely to persist so long as there are human cultures and subcultures, which is likely to mean as long as people are raised in families.

Duke University

ANTHONY APPIAH

[4] Like race-ethnicity, sexuality provides an interesting contrast case to sex-gender. It would be interesting to explore, for the sake of further contrast, the ways in which the notion of a "gay" identity requires assumptions about whether sexuality is and is not a matter of acculturation.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

STRUCTURE/ANTISTRUCTURE AND AGENCY UNDER OPPRESSION*

I shall begin this exploration into oppressed subjectivity and the subjective possibility of liberation by briefly considering two examples of theories that portray oppression as inescapable *through* the subjects' joint or separate exercise of their own volition, power, or agency. In neither case am I suggesting that the theorist does not offer also an account of liberation. But I am claiming that the account of oppression itself leaves the subject trapped inescapably in the oppressive system. The logic of the particular form of oppression leads me to understand it as inescapable.

I. TWO DESIDERATA OF OPPRESSION THEORY

Marx's alienated laborer,[1] who does not have any control over the means of production, is one example of a systematically oppressed person whose oppression is not escapable through her agency. In the logic of capitalist oppression, proletarians are not free to do anything but sell their labor power. In so doing, they enter a terrain where their labor power belongs to someone else who dictates the terms of its exercise. As labor power is used, what is produced in its consumption becomes hostile to the laborers in the form of capital. Their own productive activity becomes alien to them, since—to put it in Aristotelian terms—the syllogism that ends with the consumption of the laborers' productive activity in production is not their syllogism, but the capitalist's. The selling of labor power is the end of a syllogism of the proletarian. But this is also a peculiar syllogism, since the alternatives of the worker's choice have been manipulated. Laborers sell their labor power so as to obtain the means of subsistence that *under capitalism* they cannot obtain except through the alienation of their productive activity. The exercises that end with their alienation are the only exercises of their praxis—their conscious productive activity—open to them and these only further their alienation.

Marx also conveys clearly that proletarians must organize in the work place against capital. He gives great importance to class struggle and the production of class consciousness. But, given his account of the logic of capitalism, I do not see how that is subjectively possi-

* To be presented in an APA symposium on Gender, Race, Ethnicity: Sources of Identity, December 30. Anthony Appiah will be co-symposiast, and Thomas E. Wartenberg will comment; see this JOURNAL, this issue, 493–9 and 508–9, respectively, for their contributions.

[1] See, in particular, "Alienated Labor," in *Economic and Philosophic Manuscripts of 1844* (New York: International, 1964).

0022-362X/90/8710/500–7 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ble. I need to understand the subjective mechanisms of its possibility. If class consciousness is to be produced through necessities of particular unstable structures rather than through the subjects' exercises of their agency, such production is uninformative from the point of view advanced in this paper. As produced in part by proletarian subjectivity, class consciousness is mysterious to me. The proletarian joining with other proletarians in class struggle is not countenanced by the logic of class oppression. I do not understand the subjective mechanisms of its production. It is particularly unclear how the self can be unified and contain and express both a liberatory and an oppressed consciousness.

In her account of arrogant perception, Marilyn Frye[2] depicts the arrogation of someone else's substance for one's own projects. The arrogant perceiver in Frye's work is a man. The end of arrogant perception is the "acquisition of the service of others. The means: variations on the same theme of disintegrating an integrated human organism and grafting its substance to oneself" (*ibid.*, pp. 64–5). In successful arrogant perception,

> The intelligent body ceases to be: intelligence and bodilyness are sundered, unable to ground or defend each other or themselves. Mind and body, thus made separate, are then reconnected, but only indirectly: their interactions and communications now mediated by the man's will and interest. Mind and body can preserve themselves only by subordinating each other to him (*ibid.*, p. 66).

As depicted, the arrogantly perceived woman cannot be an agent of her own liberation. In "To Be and Be Seen" and in "Separatism and Power,"[3] Frye depicts strategies to end the very oppression that she characterizes in "In and Out of Harm's Way." But it is unclear how it is possible, given the logic of arrogation, for the arrogated woman to say 'no' and separate or how the lesbian can be in her. The lesbian as the "outsider" changes the woman's self-perception because she is being seen as someone who is not arrogated. But how can the woman see the lesbian and see the lesbian seeing her?

It is a desideratum of oppression theory that it portray oppression in its full force, as inescapable, if that is its full force. I shall not critique Marx's account of class oppression or Frye's account of sexist oppression because they depict these forms of oppression as inescapable. I think much of the explanatory power of these theories of oppression resides in their depiction of the oppressions as inescapable.

[2] *The Politics of Reality* (New York: Crossing Press, 1983).
[3] Both essays appear in *The Politics of Reality*.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

But I also consider it a desideratum of oppression theory that the theory be liberatory. The ontological or metaphysical possibility of liberation remains to be argued, explained, uncovered. If oppression theory is not liberatory, it is useless from the point of view of the oppressed person. It is discouraging, demoralizing.

So, I am recommending contradictory desiderata for oppression theory. This recommendation as well as the ontological possibility of liberation depend on embracing ontological pluralism. Ontological pluralism is suggested by many oppression theorists. Nancy Hartsock's[4] understanding of Marx as proposing different levels of reality that are both "real" suggests ontological pluralism; Frye's account of the lesbian and of the woman where the lesbian is also a woman suggests ontological pluralism; Sandra Bartky's[5] concept of "double ontological shock" and Sarah Hoagland's[6] account of femininity as resistance suggest ontological pluralism. Ontological pluralism is suggested vividly by theories of racial oppression presented by men and women of color: think of Ralph Ellison's *Invisible Man*[7] and Gloria Anzaldua's *Borderlands/La Frontera*[8] as examples.

II. THE PRACTICAL SYLLOGISM

I am understanding the practical syllogism, as Aristotle understood it in *Nicomachean Ethics,* as reasoning that ends in action, not as propositions that entail other propositions. Because I am interested in oppression, I begin by paying attention to Aristotle's account of the slave keeping the practical syllogism in mind. Aristotle makes it clear that the slave can only obey or follow orders, but cannot reason his own syllogism. The master reasons and the slave does. It is difficult to imagine the slave after working hours except as an animal: resting, eating, roaming around, copulating. But these are not human acts when performed by the after-hours slave, since they are not acts that are the end of practical syllogisms and only as such open to the kind of evaluation that Aristotle offers when he talks of practical wisdom and moral virtue.

There is an interesting similarity and dissimilarity between the slave's action in relation to the practical syllogism and the action of the arrogantly perceived woman in relation to the practical syllogism.

[4] "The Feminist Standpoint: Developing the Ground for a Specifically Feminist Historical Materialism," in *Discovering Reality,* Sandra Harding and Merrill Hintikka, eds. (Boston: Reidel, 1983).

[5] "Phenomenology of a Feminist Consciousness," in *Philosophy and Women,* Sharon Bishop and Marjorie Weinzweig, eds. (Belmont, CA: Wadsworth, 1979).

[6] "Femininity, Resistance and Sabotage," in *Femininity, Masculinity & Androgyny: A Modern Philosophical Discussion,* Mary Vetterling Braggin, ed. (Savage, MD: Roman & Littlefield, 1982).

[7] New York: Vintage, 1972.

[8] San Francisco: Spinsters/Aunt Lute, 1987.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The act of arrogating someone's substance ends the possibility of the subject's giving a practical syllogism that she can put into action and is not severely affected by the practical syllogism of the arrogant perceiver. Two practical syllogisms enter into actions performed by women who are successfully arrogantly perceived: hers and his. There is no denial here that the woman can engage in practical reasoning. But because she can, the alternatives from which she can choose and thus her conception of her well-being at the moment of action must be altered, manipulated. She then "chooses" between alternatives that she would not have chosen except for the arrogant perceiver's mediation. Once the arrogant perceiver manipulates her alternatives, she proceeds to reason practically and she chooses the alternative the arrogant perceiver wants her to choose. So, the effect is the same as in the case of the slave. But in order for there to be that coincidence, since the woman does reason practically, what he has to do is to eliminate choices that could not be "mediated by his will and interest," for example, lesbian choices, loving perceivers' choices. In a world tightly and inescapably structured by arrogant perception, lesbian reasoning, woman-loving reasoning, lacks authority or is impossible.

I am interested in the practical syllogism as explaining action or failure to act in the case of oppressed persons. I am interested in the formation of intentions, in the blocking between intention and action and in the production of what I shall call "subservient" or "servile" syllogisms, of which the syllogism of the arrogantly perceived woman is an example. I am also interested in the open-endedness of oppressed persons' acting in terms of appropriate practical syllogisms, given the context(s) or background(s) of the acting. This open-endedness is possible because the self is not unified but plural.

III. PLURALISM

I am giving up the claim that the subject is unified. Instead I am understanding each person as many. In giving up the unified self, I am guided by the experiences of bicultural people who are also victims of ethnocentric racism in a society that has one of those cultures as subordinate and the other as dominant. These cases provide me with examples of people who are very familiar with experiencing themselves as more than one: having desires, character, and personality traits that are different in one reality than in the other, and acting, enacting, animating their bodies, having thoughts, feeling the emotions, etc., in ways that are different in one reality than in the other. The practical syllogisms that they go through in one reality are not possible for them in the other, given that they are such different people in the two realities. And if one can remember the intentions of the person one is in the other world and now tries to

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

enact them, one can see that many times one cannot do so because the action does not have any meaning or a very different sort of meaning than the one it has in the other reality.

It is very important whether one remembers or not being another person in another reality. In that case, of course, one has first-person memories of that person in the other reality. There are lots of reasons why one may not remember. In the case of people who dominate others, they may not remember the persons they are in the reality of the dominated. For example, many times people act in front of their maids as if there were no one in the room. They say things and behave in ways that one can imagine as said or done only in private. When people behave this way, they do not see themselves as the maid sees them and they do not want to remember or recognize the persons who are seen by the maids, of whom the maids are witnesses. The maids can testify only in the world of the dominated, the only world where that testimony is understood and recognized. There are many reasons why the employers do not remember themselves as maids know them. One of these reasons concerns their own sense of moral integrity since, as they are witnessed by the maids, they lack it. So phenomena such as self-deception become very important in this way of seeing things in ontological pluralism.

The relation between the practical syllogism and the rejection of the 'each person is one' claim is complex. At least some of the syllogisms of the different person one is in one world will be different in kind from those in another world: they may be servile syllogisms, or syllogisms that have intentions that cannot be understood in the world in which they are being put into action. One can see that the very same person may remember herself in another reality and thus be able to form practical syllogisms that have intentions that the person she is in another reality would have. But, of course, her actions would not be visible or heard or understood in the reality in which she performs them in that case. That is the meaning of invisibility.[9] That is why it is connected to responsibility. Visibility, faithful intelligibility, are necessary conditions for responsibility. In the reality in which the person is functioning, the person is not thought of, and it is hard for that person to animate herself as a person capable of practical syllogisms that are not subservient to the syllogisms of her dominator.

So, the liberatory experience lies in this memory, on these many people one is who have intentions one understands because one is fluent in several "cultures," "worlds," realities. One understands herself in every world in which one remembers oneself. This is a

[9] See the prologue in Ellison's *Invisible Man*, esp. p. 11.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

strong sense of personal identity, politically and morally strong. The task of remembering one's other selves is a difficult liberatory task. Mystification is one of the many forms of control of our memory of our other selves. All oppressive control is violent because it attempts to erase selves that we are that are dangerous to the maintenance of domination over us.

But one may also inhabit the limen, the place in between realities, a gap "between and betwixt" universes of sense that construe social life and persons differently, an interstice from where one can most clearly stand critically toward different structures. Other worlds provide one with syllogisms that one can attempt to make actual in the worlds in which one is oppressed, given one's critical understanding of each world. The critical understanding is made possible, in part, by the going into the limen when she "travels" to the other worlds. The limen is the place where one becomes most fully aware of one's multiplicity.

My sense is that one of the factors that makes oppression inescapable in many theories that portray oppression as inescapable is the inability to form liberatory syllogisms in the world of the oppressor, given the logic of oppression. It is also my sense that the limen is a place where one can form liberatory syllogisms. But many of the syllogisms formed in the limen which would inform the fabric of everyday life cannot be completed. So one will have very good reason to attempt to understand why the syllogisms cannot be completed and attempt to correct what blocks action. So, the connection between the practical syllogism, ontological plurality, and liberatory oppression theory resides in the fact that the oppressed know themselves in realities in which they are able to form intentions that are not among the alternatives that are possible in the world in which they are brutalized and oppressed.

IV. STRUCTURE AND ANTISTRUCTURE

Victor Turner[10] uses the distinction between structure and antistructure to study pilgrimage. He borrows the term "structure" and its definition from Robert Merton, who uses the term for "patterned arrangements of role-sets, status sets and status sequences consciously recognized and regularly operative in a given society and closely bound up with legal and political norms and sanctions" (*ibid.*; p. 237).

Although structures are not closed, ordinary life presents structures as systematic, complete, coherent, closed socio-political-economic organizations or normative systems that construe persons. Structures construct or constitute persons not just in the sense of

[10] *Dramas, Fields, and Metaphors* (Ithaca: Cornell, 1974).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

giving them a facade, but also in the sense of giving them emotions, beliefs, norms, intentions. That is, the person does not just wear a mask, but that person is the person that the structure constructs. Even the experience of putting on a mask or enacting a role is structure-dependent. The husband putting on his physician's coat does not move on to another structure even if he feels that, as he is putting on the coat, he is "changing faces," changing roles. Both roles occur within the same structure, and are articulated, connected in the same structure. I am disagreeing in this respect with Turner's understanding of the person within a structure as a persona, a role mask, not "the unique individual" (*ibid.*). Turner seems to me to commit himself unnecessarily to a transcendental self that "plays the roles," puts on the masks but is distinguishable from them.

Antistructure is constituted by liminality and communitas (*ibid.*, p. 202). Liminality is a term that was introduced by the anthropologist Van Gennep in his study of rites of passage and Turner analyzes and uses it in the study of pilgrimage. It is the phase between separation and reaggregation in transition rites (*ibid.*, p. 231), the moment when the initiand is "betwixt and between successive lodgments in jural political systems" (*ibid.*, p. 13). "In this interim of liminality the possibility exists of standing aside not only from one's own social position but from all social positions and of formulating unlimited series of social arrangements" (*ibid.*, p. 14). In the liminal phase, there are no structural descriptions, so the liminal subject does not stand with respect to others in the limen as in a hierarchy. "In liminality, the symbolism almost everywhere indicates that the initiand, novice, neophyte is structurally if not physically invisible in terms of his culture's standard definitions and classifications" (*ibid.*, p. 232). The experience of victims of ethnocentric racism of moving across realities, of being different in each, and of reasoning practically differently in each can be understood as liminal. To do so is to understand its liberatory potential because, when the limen is understood as a social state, it contains both the multiplicity of the self and the possibility of structural critique. Both are important components of the subjective possibility of liberation. I think that to experience oneself in the limen is significantly to experience oneself as multiple.[11] As such, it is extremely threatening to any world or any aspect of a world that requires unification, either psychologically, morally, politically, or metaphysically.

What I find interesting and helpful in Turner is both that he describes experiences that are of a piece with the experience of

[11] Cf. *Borderlands/La Frontera.*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

victims of ethnocentric racism and that he claims that liminal states are social states, just as much as structural states are.

If we use the concept of structure and antistructure, *and* understand structure as construing real persons and not just the masks of persons, *and* we acknowledge that there is more than one structure in which the in-some-sense the same person lives, *and* we acknowledge that people can go in-between structures, *and* that it is possible to be without structure and thus without structural construction, *then* we do not have to appeal to the ahistorical, acultural, asexual, without race or gender, without a system of production within which her productive activity has meaning transcendental ego, in order to understand liberation from oppression. We also succeed in offering a noncontradictory account of the escape from inescapable oppression.

Seeing structural invisibility as an important aspect of liminality makes clear why victims of ethnocentric racism are never seen as liminal subjects. But, of course, that the liminal state is structurally invisible does not necessitate that liminal states not be sociohistorical states. The social transcends the structural without metaphysical transcendence. The historical subject is multiplicitous and sometimes liminal.

Liminality is also of a piece with my account of the multiplicitous person reasoning practically within and across worlds. But if multiplicitous persons reason within and across worlds, one need not see a person at any one point as offering or enacting or intending just any *one* practical syllogism. The practical syllogisms may be many and thus her acting may be completely ambiguous. It may be that, at the same time, she is acting in accordance with the norms of one reality, one structure, and acting with the norms of another reality, or in accordance with the possibilities that lack of structure bring up for her. This is what I see as useful for liberatory theory in Turner's insight that antistructure is social. A thoroughgoing critique of one's own situation may be a part of or inform one's practical syllogism and that may be possible because of a structural gap and because one remembers oneself in the antistructure.

But, of course, merely remembering ourselves in other worlds and coming to understand ourselves as multiplicitous is not enough for liberation: collective struggle in the reconstruction and transformation of structures is fundamental. But this collective practice is born of dialogue among multiplicitous persons who are faithful witnesses of themselves and also testify to, and uncover the multiplicity of, their oppressors and the techniques of oppression afforded by ignoring that multiplicity.

Carleton College MARIA C. LUGONES
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

COMMENTS ON APPIAH AND LUGONES*

THE title under which this symposium takes place—Gender, Race, Ethnicity: Sources of Identity—suggests a way of thinking about gender, race, and ethnicity which my predecessors have resisted, albeit only implicitly. They have suggested that we need to think of these categories not just as sources of our self-identity, but as categories by which we are classified. In my comments, I shall draw out the logic of this position more fully.

I

Anthony Appiah discusses the issue of identity in a post-Kripkean philosophic framework, distinguishing between metaphysical and ethical identity. A person's metaphysical identity has to do with the stuff from which she came, whereas her ethical identity has to do with her sense of herself, i.e., the sorts of changes that would cause her to reject her changed self as still her.

In developing his view, Appiah makes the important point that race, class, and gender are social categories, although they have been mistakenly interpreted as biological. Despite the presence of this insight, I think that Appiah's view has a number of problems. Most serious among them is the distinction between a person's metaphysical and ethical identities. Although Appiah is certainly correct to argue that the concept of metaphysical identity is not enough for social theory, I do not find his supplementation of that notion with the concept of an ethical identity to be adequate.

Let me indicate one problem I shall address: Appiah's acceptance of a serial model of identity, one in which gender, race, and ethnicity make separate and distinct contributions to a person's general sense of identity. I shall argue that a person's identity cannot be analyzed as the result of such separable contributions of distinct class, gender, racial, and ethnic identities.

II

Maria C. Lugones begins her paper with a paradox. If an oppression theory is to be successful, it must portray oppression as forceful and inescapable, as ruling the lives of those who are oppressed. It must present the identities of those who are oppressed as structured by

* Abstract of a paper to be presented in an APA symposium on Gender, Race, Ethnicity: Sources of Identity, December 30, commenting on papers by Anthony Appiah and Maria C. Lugones, this JOURNAL, this issue, 493–9 and 500–7, respectively.

0022-362X/90/8710/508–9
© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the very form of oppression it is theorizing. In so doing, however, it makes the possibility of escape from that oppression seem impossible. In order for a person even to conceive of liberation, she must have an identity that transcends the identity by which she is socially constituted as oppressed. It therefore seems that the very logic of oppression, as we have come to understand it, denies the possibility of achieving a liberatory consciousness.

Lugones's theory of memory and liminality are attempts to show that the prospects for a theory of liberation are not as bleak as they seem. By reference to the idea of competing identities within a person, Lugones hopes to replace the traditional idea that the self is the source of resistance and thus the locus of hopes of liberation.

I do not believe that Lugones has presented a way out of the dilemma she so forcefully describes. My worry is that her view points toward the very possibility that it seems to fear: a totally administered society in which the problematic borderline identities are erased.

III

I shall conclude my comments by addressing the following question: Is it possible to develop a notion of identity that can successfully mediate its two aspects: identity is that imposed from without by which others are able to categorize and objectify us, and identity is that which comes from within and allows us to constitute ourselves as desirous of liberation? I shall also ask what use the categories of gender, race, and ethnicity have in thinking about the project of social liberation.

THOMAS E. WARTENBERG

Mount Holyoke College

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE WAY OF IDEAS: A RETROSPECTIVE*

JOHN LOCKE'S *Essay concerning Human Understanding* was considered by many of his contemporaries to undermine traditional metaphysical and religious doctrines. Many years in gestation, worked over in different countries and times while serving in active roles directed more toward practical concerns than most of the *Essay's* doctrines, that work rather quickly became the object of attack. Some of the features of that work which disturbed his contemporaries were the rejection of innate ideas and principles; the apparent denial of substance as a basic metaphysical category (if not a denial, an indifference); the stress upon experience and observation, which seemed to reject rational principles and which was hesitant about theory; the skepticism about nonobservable features of reality, such as real essence; his suggestion that God could superadd thought as a property to matter (e.g., the brain); and the very language of ideas.

The reaction to the language of ideas is rather surprising since Locke did not invent it: he took it over from the Cartesians, especially as Antoine Arnauld interpreted the term 'idea' in Descartes. Many writers reacted to that term in the *Essay* as if it were a "new language." Henry Lee, Peter Browne, John Milner, and others joined Bishop Stillingfleet[1] in finding the language of the way of ideas difficult and obstructionist. Stillingfleet recognized an ordinary use of the term 'idea', but he charged Locke with using that term in a technical way which Stillingfleet found objectionable.

> But your Way of *Certainty by Ideas* is so wholly New, that here we have no *general Principles;* no *Antecedents and Consequents;* no *Syllogistical Methods* of Demonstration; and yet we are told of a better Way of Certainty to be attained, meerly by the help of *Ideas* (*ibid.*, p. 120).

Despite the widespread objection to this "new" language, many of the books coming after Locke, even those attacking him, began to employ the same language, even some of the same doctrines. The way of ideas became influential and infamous almost at the same time.

* To be presented in an APA symposium on John Locke after Three-hundred Years, December 30. James Tully will be co-symposiast, and Vere Chappell will comment; see this JOURNAL, this issue, 517–523 and 524–5, respectively, for their contributions.

[1] *The Bishop of Worcester's Answer to Mr. Locke's Second Letter* (1698). For the details on the reaction to the language of ideas in the *Essay*, see my *John Locke and the Way of Ideas* (New York: Oxford, 1956), pp. 86–92.

0022-362X/90/8710/510–6
© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The language of ideas employed by Locke set the style for much subsequent writing. One of the most active topics among the Cartesians preceding Locke (e.g., Malebranche, Arnauld) was that of ideas: their role in knowledge, their nature, their origin. Locke borrowed and refurbished the Cartesian 'idea', particularly as interpreted by Arnauld against Malebranche. The question of realism—direct or indirect—given prominence by the extended Arnauld-Malebranche debate in the 1680s (a debate well-known to Locke) was joined by Locke's use of that term. 'Idea' for Locke was not a term he felt the need to clarify and define precisely. Passages from his *Essay* can be found which sound like Arnauld or like Malebranche. I think we can discern where Locke stood on the issue—"Are ideas special entities?"—but that question is still under debate today.

I

It is difficult to talk about the *Essay* in the last decade of the seventeenth and the first decade of the eighteenth century apart from Locke's other major writings: *Two Treatises of Government, Some Thoughts concerning Education, The Reasonableness of Christianity,* the works on toleration, even some of the tracts on money. Locke's popularity was a function of this large body of writing. It is also important to recognize that he was known for his roles in government after 1688 and his position as secretary to the Earl of Shaftesbury prior to that date (with the trouble that association brought). Locke was not just a writer, albeit a writer on a wide range of topics. He was a friend of scientists, medical doctors, and government persons. Deprived of his membership in Christ Church by one king, Locke was recognized and rewarded by another. Parents sought his advice on raising children, friends frequently wrote him for medical advice on various illnesses. The popularity of the *Essay,* even some of the attacks on it, may have been due to the prominence that his other books and his wider activities gave him.

The early years of the eighteenth century in Britain saw some of the best criticisms and defenses of the *Essay:* writers such as Milner, Norris, Cockburn, Lee, Broughton, Proast, Bold. The *Essay* also influenced a number of writings on logic, e.g., by Duncan and Watts, and it was reflected in many other books on logic and metaphysics. The entry on "metaphysics" in the first edition of the *Encyclopaedia Britannica* consisted of long passages from the *Essay.* Chambers's 1728 *Cyclopaedia* borrowed material from Locke for a number of entries. As in the seventeenth, so in the eighteenth century, the doctrines of the *Essay* gave rise to disputes, the main one in the latter century being over thinking matter. In Britain alone, there was an extended debate about Locke's suggestion that God could superadd

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

to matter the power of thought, a debate involving scientists, theologians, medical men, and literary figures.[2] In France, that same controversy took its own form, spearheaded by Voltaire's Letter XIII in his *Letters concerning the English Nation* (1733).[3] Journal articles, book reviews, many clandestine tracts discussed and defended Locke's suggestion. Reinforced by Condillac's claim that he was extending Locke's philosophy, the thinking-matter suggestion combined with what was called Locke's "sensualism" (and the causal theory of perception) to make Locke one of the main figures for those who opposed French materialism. By mid-century, the revival of Malebranche in France began to make inroads in Locke's popularity, despite or perhaps because of Diderot's warm, praiseworthy entry on Locke in the *Encyclopédie*. That particular controversy had died down by the end of the eighteenth century. Locke's name all but disappears from journal discussions by the 1760s. In England, the year 1764 saw Thomas Reid's *An Inquiry into the Human Mind,* followed in 1785 by his *Essays on the Intellectual Powers of Man,* both of which but especially the latter gave prominence to Locke's doctrine of ideas. In Germany, Leibniz's *Nouveaux Essais* appeared in 1765. Each in their own way, Reid and Leibniz set the scene for subsequent discussions of topics in Locke's *Essay*.

The early nineteenth-century French-language historians, headed by Victor Cousin, sealed Locke's fate as the leader of the school of "sensualism." Only two strong challenges to that reading of Locke appeared in that century: Dugald Stewart's 1810 "On the Influence of Locke's Authority upon the Philosophical Systems Which Prevailed in France," and Thomas Webb's spirited attack on Cousin's 1834 *Cours de philosophie* in his *The Intellectualism of Locke* (1857). The number of articles on Locke began to increase in the last years of the nineteenth century. Those authors writing on Locke came from different countries.

In our own century, the articles and books on Locke took a diversified shape. If we consult some of the standard journals of the first forty years of our century, we will find discussions of most of the main doctrines in the *Essay*. German and Italian scholars became very active on most aspects of that work, as well as on other writings by Locke. Biographical books and articles appeared, some editions of correspondence. Alexander Campbell Fraser, whose 1894 edition of the *Essay* was to exert such an influence on English-speaking readers

[2] For the details of this debate, see my *Thinking Matter: Materialism in Eighteenth-Century Britain* (Minneapolis: Minnesota UP, 1983).

[3] For this fascinating story, see my *Locke and French Materialism* (New York: Oxford, forthcoming).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of Locke, contributed the entry on Locke in the eleventh edition of the *Encyclopaedia Britannica* (1910). From that year to the 1940s, publications on Locke blossomed and proliferated. To attempt a listing of the important English-language contributors in these years would be perhaps subjective, but one must mention such names as James Gibson, Sterling Lamprecht, John Dewey, George Santayana, Reginald Jackson, Christopherson (author of the first systematic bibliographical study of Locke), Richard Aaron.[4]

The years 1949–50 mark an important turning point in Locke studies with the acquisition of the Lovelace collection of Locke manuscripts, drafts of the *Essay*, journals, notebooks, letters. Those papers enabled scholars to study the development of Locke's thought, to follow his book purchases and the growth of his library, to appreciate the international scope of his correspondence. One might almost say that the serious scholarly study of Locke waited upon the public access to the Lovelace collection and the remains of Locke's library, both of which are now in the Bodleian Library at Oxford. The magisterial eight-volume edition of Locke's correspondence compiled over many years by E.S. de Beer, and the edition of the *Essay* by the late Peter Nidditch, signaled the decision by the Oxford University Press to publish for the first time since 1824 the works of Locke. At this date, we have published in that series besides the correspondence and the *Essay*, the *Paraphrases of St. Paul's Epistles* and *Some Thoughts concerning Education*. Three volumes of drafts of the *Essay*, with other related papers, will begin to appear over the next few years. The important journals (four or five volumes) will also soon be published. Locke's writings on money are now in press. Other Locke titles are in various states of preparation. The time may not be far off for a new biography.

II

In all this very recent activity, the *Essay* and its doctrines play a central role, although *Two Treatises* has had at least equal time. The *Essay* is such a large and complex work that it is difficult to try to identify all the issues and questions that are raised and discussed there. But we can draw up a list of central themes, themes seen as important by his seventeenth-century readers and by philosophers today. Like Hume's *Treatise*, Locke's *Essay* is in many ways a compendium of topics then in vogue. Both authors also challenged and broke away from a number of traditional concepts. At the risk of over simplification, I would characterize the *Essay* in terms of one

[4] For a full bibliographical listing and description of books and articles on Locke from 1689 to 1982, see Jean S. and John W. Yolton, *John Locke: A Reference Guide* (Boston: Hall, 1985).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

central concept, substance, and the methodology it recommends, experience and observation.

Bishop Stillingfleet charged Locke with eliminating the category of substance altogether from his ontology. The Bishop had his eye upon the theological doctrine of the Trinity (he never succeeded in eliciting from Locke a clear response for or against that doctrine), but he, like many others, was also reacting to what appeared to be an indifference on Locke's part to the traditional dualism of two substances. Locke is not unequivocal on this topic. He provides us with a descriptive account of how and why we feel the need to find a subject for the experienced qualities that we group into kinds, but at the same time he shows us (as did Hume more decisively later) that we have no basis for a knowledge claim about substances over and above qualities. On the lack of a knowledge of substance, Locke was just as firm about material as he was about immaterial substance, at least for the substances accepted by philosophers and theologians. Locke does appeal to a theory that he may have seen as a replacement for the vague metaphysical "something I know not what": his appeal to the corpuscular theory of matter generally accepted by Boyle and other scientists of his day. This was a theory that went beyond the methodology of experience and observation, with its talk of insensible particles. Locke accepted this theory as the best explanation of observed phenomena. A case can be made for saying that Locke's real essence is identified with the corpuscular structure of the matter of objects. We cannot go so far as to say that corpuscular matter plays the same role for Locke as did metaphysical substance for traditionalists, since, among other reasons, we cannot predicate all observed qualities of the corpuscular structure without expanding on a suggestion Locke (and Boyle) makes about dispositional properties. Nevertheless, the structure *was* in Locke's time unobservable and it *was* used to explain phenomena; at least Boyle offered some explanations of this sort. Locke invoked that structure to account for the causation of observed qualities and perhaps (there are serious difficulties here) also for perception. It could be said that Locke replaces the substance metaphor of 'standing under' by the relation of 'caused by', a replacement of metaphysics by science.

There is a kind of counterpart to this analysis of matter in Locke's radical concept of personal identity. We cannot take that concept as applying to all spirits (i.e., immaterial substances in the older vocabulary, e.g., angels or God), but what worried some of Locke's more traditional opponents about his account of personal identity was the absence of any appeal to an immaterial substance. In fact, even the identity of plants and animals is located in the continuous life of

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

those individuals, not in any underlying substance. With persons, sameness of soul, 'soul' being the traditional substance term, was on Locke's analysis not explicitly denied but it was deemed irrelevant to the sameness of persons. It was sameness of consciousness that carried the identity of a person. Unlike the corpuscular structure of matter which was not experienced, consciousness accompanies every mode of awareness and is a feature of our experience. Material objects for Locke are individuated by time, place, and corpuscular structure, persons (nonmaterial components of reality) are individuated by time, place, and continuity of consciousness.

The traditional substance-quality metaphysics was challenged in another way by Locke, his suggestion that a property such as thought, which was always considered the prerogative of immaterial substance, might be made a property of matter. Such a suggestion, quite apart from its overtones of materialism, upset the standard view of the autonomy of the two substances, each with its essential, nontransferable properties. Especially in eighteenth-century France, Locke was associated with the system of physical influence, the notion that matter does have causal relations with mind. Locke admitted that he did not understand how ideas could be caused by particles in motion activating the physiology of the body, but he nowhere explicitly denied such influence. For a variety of reasons, occasionalism and pre-established harmony became the dominant accounts (really denials) of mind-body (or at least body-mind) interaction. One possible alternative to all three of these systems was some form of materialism, where thought and its modes become properties of the brain. Locke did not espouse this radical view but he did suggest its possibility, albeit only under God's action. Such a view would end with a dualism of properties but a monism of substances—if indeed, a functioning brain were (as I think it was not by Locke) viewed as a material substance. So once again, Locke was at odds with traditional metaphysics, even though his suggestion was speculative and not an asserted doctrine.

Still another disturbance to long-held beliefs was Locke's apparent rejection of *natural kinds*, at least his firm assertion that our classifications are conventional, man-made, and dependent upon observed qualities. There *are* those many passages where Locke details the a priori knowledge we would have were we able to penetrate to and acquire an understanding of the real essence, *his* real corpuscular essence. His acceptance of the great chain of being, with no chasms or gaps, his apparent belief in monstrous births, would seem to reinforce a rejection of natural kinds, although he does not write enough about biological species to enable us to say with confidence

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that a scale of being dissolves boundaries between species. So I think it is not clear just what those extended discussions of a possible deductive science of nature entailed for Locke. One conclusion does seem clear: our knowledge does not extend to such insights into the corpuscular structure, *our* classifications in terms of observed qualities work well enough for physical science. The *Essay* is, I believe, silent on any implications for biology.

III

The most recent interest in Locke's *Essay,* as with some other books on perception in the seventeenth and eighteenth centuries, lies in the key term in this way-of-ideas tradition, the term 'idea'. Contemporary discussions of realism and representation have begun to recognize some similarities between their accounts and Descartes, Arnauld, and Locke on ideas and representation. The psychology of these older writings contain anticipations of very recent notions. A historical tracing of this aspect of Locke and the way of ideas from Descartes through Reid and Kant can, I suspect, shed light on recent debates. Such a tracing will at least remind us of the history of these concerns; it may even offer some resolutions to questions about representation and realism. We need to decide what Locke's ideas were, how he understood their cognitive content to yield knowledge of the physical world, whether the usual understanding of the representative theory of perception was in fact held by Locke, whether, as Arnauld insisted against Malebranche, we can have representation and realism with ideas. Locke's *Essay,* when placed in its context of writings on knowledge and perception, can still be a fruitful source of information and analysis. In most of its doctrines, the *Essay* continues to stimulate some philosophers, frustrate others, and attract intellectual historians by its central and pervasive role in seventeenth- and eighteenth-century debates.

Three hundred years on, the *Essay* may not stir quite the emotional reactions it did in those centuries, but I believe it is still very much alive, readable, and relevant. It also continues to fetch high prices at antiquarian book auctions, as a recent event in New York City confirmed. I do not, however, despite the outrageous prices a draft and first edition of the *Essay* fetched at that auction, believe its value as a collectible has outstripped its intellectual worth.

JOHN W. YOLTON

Rutgers University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

POLITICAL FREEDOM*

THREE hundred years after the publication of the *Two Treatises of Government* and *A letter concerning Toleration,* one aspect of John Locke's thought continues to stand out in sharp contrast to the prevailing conventions of both seventeenth- and twentieth-century political theory. This is his hypothesis that institutionalized forms of government are derived from and perpetually rest upon the prior freedom of the people to exercise political power themselves (summarized at 2.171).[1] Locke had no doubt that it would "seem a very strange Doctrine to some" in his own time (2.9). It is no less unconventional today. In the current debate between liberals and civic humanists, for example, the leading participants share the assumption that political freedom is derived from and rests upon basic institutions and traditions.[2] The aim of this paper is to enable readers to understand Locke's strange doctrine by drawing attention to its distinctive features, first by means of a synopsis and then by a series of contrasts with more conventional views.

I

Prior to the establishment of institutionalized forms of government, people are capable of exercising political power themselves in an ad hoc manner: "the *Execution* of the Law of Nature is in that State, put into every Man's hands" (2.7). The exercise of political power comprises the abilities to know and to interpret standards of right (natural laws), to judge controversies concerning oneself and others in accordance with these laws, and to execute such judgments by punishments proportionate to the transgression and appropriate for purposes of restraint and reparation (2.7–12). Individuals are free to order their actions within the bounds of natural laws and are equal in the "Power and Jurisdiction" to govern the actions of those who transgress these bounds (2.4, 2.6).

Since "every one is Judge, Interpreter, and Executioner," this ad

*To be presented in an APA symposium on John Locke after Three-hundred Years, December 30. John W. Yolton will be co-symposiast, and Vere Chappell will comment; see this JOURNAL, this issue, 510–6 and 524–5, respectively, for their contributions.

[1] The numbers in parentheses throughout refer to treatise and section numbers of Locke, *Two Treatises of Government,* Peter Laslett, ed. (New York: Cambridge, 1970).

[2] John Rawls, "Justice as Fairness: Political Not Metaphysical," *Philosophy and Public Affairs,* XIV (1985): 223–252; Richard Rorty, "The Priority of Democracy to Philosophy," in *The Virginia Statute of Religious Freedom,* M. Peterson and R. Vaughan, eds. (New York: Cambridge, 1987); Michael Sandel, *Liberalism and the Limits of Justice* (New York: Cambridge, 1982); Charles Taylor, *Philosophical Papers, Vol. 2* (New York: Cambridge, 1985), pp. 185–339.

0022-362X/90/8710/517–523

© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

hoc form of self-government, or individual popular sovereignty, does not run smoothly in practice (2.136). People disagree on a "common measure to decide all Controversies between them" and they are reluctant to concede that a law is binding in their own case (2.124). They lack the motivation to participate in the ad hoc adjudication and prosecution of disputes when the case does not concern them, and they tend to participate with passion and revenge when it does (2.125–6).

To overcome these difficulties, people set up institutionalized forms of government by uniting into a community and conditionally placing their political powers in the hands of designated legislators and executives (2.87–9, 95, 132, 144). Accordingly, political communities and governments not only derive from, but also perpetually rest on, the abilities of the citizens to judge and act politically. The members entrust their political power on the condition that the governors will exercise it in accordance with the public good. If the people find that their trustees (either the executive or legislators) abuse the power given to them and "act contrary to the trust reposed in them," the bond of obligation between governors and governed is forfeited, political power devolves "into the hands of those that gave it," and they may use it to remove them. "And thus," Locke emphasizes, "the *Community* perpetually retains a *Supream Power* of saving themselves from the attempts and designs of any Body, even of their Legislators, whenever they shall be so foolish, or so wicked, as to lay and carry on designs against the Liberties and Properties of the Subject" (2.149). As he explains in more detail in the final chapter, when political power devolves to the people, they exercise it in the form of a revolution to remove their untrustworthy governors, just as they would proceed against lawbreakers prior to the establishment of legislative and executive institutions, and "the People have a Right to act as Supreme, and continue the Legislative in themselves, or erect a new Form, or under the old form place it in new hands, as they think good" (2.243).

Who has the right to judge when the trust has been forfeited and government dissolved? Locke's answer is that "*The People shall be Judge*" and by this he means "*every Man* is *Judge* for himself" (2.240–1). Then, the "Body of the *People*" should execute this judgment by forcefully removing the unjust rulers. If this fails, each person who so judges has the right to act as best one can (2.242–3; cf. 2.168). Locke stresses that "*this Doctrine* of a Power in the People" will not lay "a *ferment* for frequent rebellion" (2.224–230) because people will not actually engage in a revolutionary contest until oppression touches the majority, or touches only a few but seems to threaten all (2.208–210, 2.230, 2.168).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The final claim Locke advances is that his account, in addition to being an explanation and justification of resistance to oppression, is also the best means of hindering oppression in the first place. Under any institutionalized form of government, those in power will be tempted in various ways to abuse the power entrusted to them: to develop an interest separate from and in violation of the good of the people (2.209–210, 2.228). This theme is emphasized even more emphatically in *A Letter concerning Toleration*[3] where he analyzes the post-Reformation case of government imposing a uniform religion and persecuting religious dissenters. The "properest way to prevent the evil, is to show them the danger and injustice of it"; that is, to confront power holders with a citizenry who know the popular basis of government, who constantly judge the public actions of their trustees, and who are always ready and willing to exercise political power themselves when it is abused (2.226; cf. 2.228).

II

The distinctive features of Locke's portrayal of political freedom can now be thrown into clear relief by contrasting them with a number of seventeenth-century alternatives. First, the conventional means to check the abuse of institutionalized forms of political power are all found by Locke to be insufficient. The argument, advanced by proponents of mixed and balanced government, that representative institutions, the division of powers, and the rule of law constitute sufficient checks was falsified in practice by the ability of seventeenth-century absolute monarchs to override these limitations (2.107, 2.111–2, 2.163). Further, these institutional arrangements, while important, are not always sufficient to stop an oppressive representative body (2.138, 2.149, 2.221–2). Locke is skeptical of the republican belief that participation in government fosters citizens oriented to the public good and impervious to the temptations of power (2.156, 2.223). He also rejects the claim that, as long as citizens are free to appeal to courts and parliaments, they will be able to hinder oppression and thus render revolution unnecessary. Locke learned from the unsuccessful attempts by religious Dissenters in the 1660s and 1670s to gain civil and religious liberties by legitimate means that those in power could block the Dissenters' appeals, stigmatize them as "seditious," and introduce repressive legislation (2.218). In *A Letter concerning Toleration,* he argues that the separation of church and state, and a clear distinction in law between politics and religion, along with representative institutions and a Bill of Rights, while necessary, are not sufficient to defend a free and tolerant society from oppression. The ultimate guardian is again said

[3] James Tully, ed. & intro. (Indianapolis: Hackett, 1983).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

to be the ability of the people to judge if their governors are ruling in accordance with the public good and to be ready to remove them if they are not (*ibid.*, pp. 52–5).

Second, Locke's location of political power in the people contrasts sharply with the customary beliefs of his contemporaries. The rebellions of early modern Europe were standardly understood to be nonpolitical acts of self-defense by the people against a ruler who had attacked them. They were not thought of as political acts, because, on the conventional view, political power was never originally in the hands of the people and, consequently, could not devolve back to them. In addition, the "people" was interpreted as either a single person acting in self-defense or an established representative of the people (usually a parliament or an estate) defending them against an attack by a monarch. Theorists of mixed and balanced government could not claim that the representative bodies, let alone the people, exercised political power in a rebellion without undermining their theory that political power always exists in a mix or balance between monarch and representative bodies. Republicans could not say that the people exercise political judgment and power outside of political institutions, because they held that political judgment and power come into being with the establishment of an authoritative political community. And all these authors agreed that the worst conceivable basis for government would be the judgment of the multitude.

These commonplaces began to appear questionable in the light of the English Civil War. In the latter phase, the people took up arms against their own representatives and they presupposed the right to exercise political power, not just to defend themselves, in trying and executing Charles I and in setting up a new form of government. Either one held to the conventional understanding, and thereby deemed this exercise of popular sovereignty illegitimate, as Kant was classically to do, or one reconceptualized the power of the people, and thereby deemed this judicial proceeding legitimate, as Locke did. Furthermore, Locke wrote and published the *Two Treatises* to justify revolt against the persecution of religious Dissenters during the Restoration. Since the Dissenters were excluded from public office and lacked sufficient support in Parliament, he needed to ground the right to judge, initiate revolt, and reform government in the people themselves. The extent to which Locke deviated from orthodoxy can be measured by the widespread repudiation of the *Two Treatises* in 1690 because it presented an unacceptably radical interpretation of the Glorious Revolution.[4]

[4] See Julian Franklin, *John Locke and the Theory of Sovereignty* (New York: Cambridge, 1978); Richard Ashcraft, *Revolutionary Politics and Locke's Two Treatises of Government* (Princeton: University Press, 1986); Mark Goldie, "The

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

III

A contrast with the current debate between liberals and civic humanists can now be used to complete this survey of Locke's understanding of political freedom. Both sides start from the assumption that the foundation of a modern democratic society is the basic structure of post-Reformation institutions of representative government, the rule of law, religious toleration, and civil liberty. Accordingly, the role of the political philosopher, as John Rawls typically states, is to explicate or articulate the "basic intuitive ideas that are embedded in the political institutions of a constitutional democratic regime and the public traditions of their interpretation" (*op. cit.*, p. 225). This consists in making explicit the implicit consensus on conceptions of political agency and social cooperation which underlie surface disagreements and accord with the basic institutions. Despite their disagreement on what forms of political agency and cooperation best serve to support and enhance a democratic society, civil humanists are just as concerned to grant priority to political institutions, practices, and communities. Political institutions and public traditions of their interpretation, one might say, are taken to be sovereign.

The leading writers in the debate have not considered Locke's opposite hypothesis that political institutions and traditions rest upon the political freedom and popular sovereignty of the people. Instead, they have misinterpreted Locke from within their very different conceptual framework and, as a result, assimilated him to either Hobbes or Benjamin Constant, both of whom share their premise of institutional sovereignty. With Locke's hypothesis elegantly disposed of in this preemptive fashion, the conventional institutional theories, which Locke found insufficient to explain how concentrated political power can be held in check, have been recirculated as the solutions to contemporary problems.

This anti-Lockean consensus can be partly explained by the theoretical sources of the debate. Liberals draw on Kant, who lays it down that the basic structure of law and its agents is sovereign, not the consent of the governed, and can never rightfully be disobeyed.[5] Civic humanists draw on Aristotle, Cicero, Machiavelli, Rousseau, and Hegel, all of whom concur that the necessary precondition of

---

Roots of True Whiggism 1688–94," *History of Political Thought*, I, 2 (1980): 195–236; Mark Goldie, "The Revolution of 1689 and the Structure of Political Argument," *Bulletin of Research in the Humanities*, LXXXIII (1980): 473–564; and my "John Locke," *The Cambridge History of Political Thought: 1450–1700*, James Burns, ed. (New York: Cambridge, 1990).

[5] John R. Wallach, "Liberals, Communitarians, and the Tasks of Political Theory," *Political Theory*, XV, 4 (November 1987): 581–611, esp. pp. 587–8.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

and limitation on political judgment and action is an established and authoritative political community. Moreover, ever since the Reaction to the French Revolution and the attack on Lockean popular sovereignty by Burke and Bentham, there has been a widespread tendency to take political, social, and economic institutions as foundational and then to ask what forms of freedom are compatible with them. Constant,[6] for example, set forth this picture of the relation between freedom and institutions in a form that has become the "horizon" of modern political thought for many liberal thinkers. Revolution as well has been reconceptualized as a social movement caused by the underlying social and economic institutions of the modern world, rather than as a political action in response to misgovernment.

The main theoretical argument against Locke is that the abilities to judge and act politically are acquired through practice, and practice presupposes and takes place in the context of practices—political institutions and communities. Therefore, to rest these institutions on the prior abilities of the people to judge and act politically is a philosophical mistake as well as a recipe for anarchy and confusion in practice.

Since Locke wrote extensively on how abilities to judge and act are acquired by practice in practices, it is difficult to believe that he overlooked this point when he came to write the *Two Treatises*.[7] Rather, the mistake may lie with his opponents. Their assumption appears to be that political abilities are acquired in *canonical* political institutions: namely, exactly those political institutions occupied by members whose overlapping consensus of judgments and actions is said to be authoritative. So, the judgments and actions of those who are outside these institutions, or are critical of them, will be, by definition, unreliable and illegitimate. Locke was, of course, familiar with this form of argument, used for example by Anglican royalists against Dissenters, and he devoted most of book one of *An Essay concerning Human Understanding*[8] to exposing its self-validating circularity.

The argument is mistaken because people are able to acquire and exercise political abilities outside the canonical institutions and traditions. Indeed, they often acquire them through the experience of

[6] "The Liberty of the Ancients Compared with That of the Moderns," in *Political Writings*, Biancamaria Fontana, ed. (New York: Cambridge, 1988).

[7] See my "Governing Conduct," in *Conscience and Casuistry in Early Modern Europe*, E. Leites, ed. (New York: Cambridge, 1988); and John Dunn, "Bright Enough for All Our Purposes: John Locke's Conception of A Civilized Society," *Notes and Records of the Royal Society in London*, XLIII (1989): 133–153.

[8] Peter Nidditch, ed. (New York: Oxford, 1975), 1.3.20–2; cf. 1.4.22–6. Rorty has carried this form of argument further in "Thugs and Theorists," *Political Theory*, XV, 4 (November 1987): 564–580.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

exclusion from and oppression by these institutions. The Dissenters' struggle for religious toleration and political enfranchisement and the European women's movements are well-known examples of this. Further, according to Locke, the faculty of judgment, even though it is always developed in the context of a practice, involves the reflexive ability to suspend and examine the authoritative traditions of that practice itself.[9] Finally, unlike Locke's hypothesis, this form of argument cannot explain how political institutions are set up or changed. Either they must be taken for granted, as in the current debate, or a mythical legislator is required, as in the republican tradition.

The major pragmatic argument against Locke is that members of modern societies are shaped and constituted by the massive and complex institutions of power and authority. Therefore, from a sociological point of view, their judgments and actions derive from and rest upon these institutions rather than vice versa. The absence of revolutions in Europe since 1850 is often said to support this and to falsify the theory of popular sovereignty.

In reply, one might point to the remarkable revolutions that occurred in Eastern European countries over last winter. These outbreaks of popular sovereignty were neither predicted by, nor did they conform to, the social and economic theories of revolution available in the social sciences. The people involved did not have the background representative institutions that are said to be necessary to foster political freedom and judgment. Nor did they have the experience of participating in republican institutions in order to develop civic virtue and discern the common good. Their activity of civil dissent prior to the revolutions was also insufficient to reform their governments. Yet, for all that, their revolts were not simply acts of self-defense. The people claimed to respond to political oppression by governments whom they judged to have forfeited their trust (often in these terms). They proceeded to overthrow their governors, to put them on trial and punish them, and went on to set up new forms of government and new governors, as they thought good.

McGill University JAMES TULLY

[9] For one way of working out this line of argument, see my "Wittgenstein and Political Philosophy," *Political Theory*, XVII, 2 (May 1989): 172–204.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

LOCKE'S MORAL PSYCHOLOGY*

MY fellow symposiasts have produced elegant papers, though on quite different subjects; and for different reasons I have little to add to or criticize in them. So I shall not actually serve as a commentator in this symposium. What I shall do is explore on my own some of the territory that lies between Locke's epistemology, treated by him in the *Essay*, and his political philosophy, set forth in the *Two Treatises*. The territory I mean is Locke's theory of ethics. Specifically, I plan to focus on Locke's moral psychology, his account of the thinking and willing we do in our efforts to fare well in our lives, to act rightly and achieve satisfaction and virtue.

The main question Locke considers is this: What determines the will of a man who cares about what he does? Locke answered this question at length in the chapter "Of Power" in the *Essay*. But in the second edition of this work, he repudiated what he had written in the first, and provided a different answer, spelled out at even greater length. Furthermore, he kept tinkering with this new doctrine of motivation for the rest of his life, making significant changes in the text of the chapter for both the fourth and the fifth editions of the *Essay*. He also developed his position in a remarkable correspondence with van Limborch which carried into his final years.

In my paper, I lay out Locke's two doctrines of motivation and note their differences. Initially Locke says that it is a man's judgment as to what is best which determines his will. Furthermore, once the judgment is made, his willing in accordance with it follows inevitably: there is no freedom in willing (though agents often have freedom to act once they have willed).

In his later account, Locke says that it is not one's judgment of goodness but a feeling of "uneasiness" that determines the will. He also introduces a doctrine of suspension, according to which a deliberating agent may stop her will from being determined by any feeling, so that she may examine more fully the implications of the various actions that are open to her.

But does not this power of suspending one's willing amount to a measure of freedom with respect to it? This question gave Locke

* Abstract of a paper to be presented in an APA symposium on John Locke after Three-hundred Years, December 30, commenting on papers by John W. Yolton and James Tully, this JOURNAL, this issue, 510–6 and 517–523, respectively.

0022-362X/90/8710/524–5
© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

some trouble, but he seems eventually, with the help of van Limborch, to have worked out the right answer to it. Locke and van Limborch continued to differ, however, on the precise nature and role of judgment in willing; and Locke kept his distance from the Arminianism of his Dutch friend.

At the end of my paper, I try to spell out Locke's conception of rationality in willing and acting, and the relation between rationality and moral rectitude.

VERE CHAPPELL

University of Massachusetts

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

PRAGMATIC NATURALISM: THE PHILOSOPHY OF SIDNEY HOOK (1902–1989)*

THE philosophical career of Sidney Hook spans nearly a century of ideological confrontation. Deeply involved in the great intellectual and social controversies of his time, Hook's role is notable in that he was not interested in purely theoretical issues for their own sake, but always sought to relate his philosophical reflections to concrete praxis. He was, in one sense, the American equivalent of Socrates, a gadfly who was embroiled in the public arena, goading us to think clearly about our basic conceptions and values, to define what we meant, and to justify our convictions. In this Socratic endeavor, he often left no one unscathed. His probing analytic mind attempted to expose sham and inconsistencies, to judge claims by their consequences. But unlike Socrates, he used his dialectical skills to defend democracy and humanism. A masterful polemicist, he often infuriated both left and right in his quest for the truth as he saw it.

Early in his career, Hook made his mark as a leading American Marxist scholar. A youthful proponent of the revolutionary Marx in his books, *Towards the Understanding of Karl Marx*[1] and *From Hegel to Marx,*[2] he soon became disenchanted with Marxism, and in later works, *Reason, Social Myths and Democracy*[3] and *The Hero in History,*[4] he re-examined and modified his earlier Marxist assumptions. Hook's mature position in political philosophy emphasized democratic socialism, which, he thought, the Leninist interpretation of Marx had abandoned; and in his later life he called himself a social democrat, to emphasize his primary commitment to democracy. His writings were strikingly prophetic—for the dramatic changes in Communist ideology which have since occurred seem to echo Hook's earlier critiques. In a real sense, his philosophical indictment of totalitarianism has been vindicated by history.

Hook was one of the chief heirs to the pragmatic and naturalistic traditions in American philosophy, though unlike many other pragmatists he did not simply philosophize about what pragmatism was, but sought to apply its principles by practicing them. In this regard,

* To be presented in an APA symposium on the Philosophy of Sidney Hook, December 30. Nicholas Capaldi will comment; see this JOURNAL, this issue, 535–6.

[1] London: Golancz, 1933.
[2] New York: Day, 1936.
[3] New York: Day, 1940.
[4] New York: Day, 1943.

0022-362X/90/8710/526–534
© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

he demonstrated the full application of the pragmatic temper as it is exemplified in practical conduct.

I. THE NATURE OF PHILOSOPHY

Hook's conception of philosophy is related to both his pragmatic naturalism and the influence of Marxism on his thought. With the pragmatists, he maintained that ideas are "implicit plans of action," which have bearing on the affairs of experience.[5] With Marx, he maintained, "practice is the life of theory; theory the guide to practice. The subject matter of philosophy is . . . the specific problems of man, history, and culture."[6] Philosophy is in this sense a prospective anticipation of the future. Like John Dewey and Marx, Hook maintained that philosophy had a role to play in civilization. He insisted that "philosophy is legitimately concerned with [the] large problems of human affairs, [and] that philosophers should have something to say to their fellow citizens and not only to their professional colleagues."[7]

What is distinctive of the philosophical enterprise, he thought, was that it was concerned with the nature of value and the good in human beings, society, and history. What do Descartes, Leibniz, Hobbes, Spinoza, and others have in common, he asked? The most comprehensive conception of philosophy that emerges from the history of philosophy, says Hook, "is that it is the normative consideration of human values." If this is as philosophy has been or if it is taken as a proposal of what it should be, it is only to be expected that the community should turn to the philosopher to find some "wise guidance" or "integrated outlook on life" (*ibid.*, p. 6). Whenever a society seeks to develop a reflective morality, it needs to clarify the structure and the role of reason and knowledge in the conduct of life, in virtue of which it can perform its function, and the constitution of nature and the universe in terms of which knowledge and the good can be reached.

II. HOOK'S THEORY OF KNOWLEDGE

Hook described his philosophy as pragmatic or experimental naturalism. He thought that the term 'experimental' was appropriate in the sense that it emphasized the fact that ideas must be tested by their experimental consequences. But he cautioned that experimentalism did not imply any kind of universal skepticism; for we are able

[5] *The Quest for Being and Other Studies in Naturalism and Humanism* (New York: St. Martin's, 1961), p. 4.
[6] "New Introduction," in *From Hegel to Marx: Studies in the Intellectual Development of Karl Marx* (Ann Arbor: Michigan UP, 1962), p. 24.
[7] *Philosophy and Public Policy* (Carbondale: Southern Illinois UP, 1980), p. 5.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

to derive reliable knowledge about the world and human affairs. For Hook, pragmatism is a philosophy of life that "holds that the logic and ethics of scientific method can and should be applied to human affairs." Unlike ethical skeptics, Hook held that one can frame warranted assertions about values as well as facts and "that it is possible to gain objective knowledge not only about the best means available to achieve given ends . . . but also about the best ends in the problematic situation in which the ends are disputed or become objects of conflict."[8] Hook thus believed that there are some criteria of objectivity that can serve as an ideal in testing truth claims. "Unless there is some appeal to evidence or argument that has some universal validity—call it reason or science or common sense—there is no more ground for accepting these criticisms than for rejecting them" (*ibid.*). Here Hook, like Dewey, is interpreting 'scientific' in a broad, not narrow, sense. Hook agreed with Dewey's statement that 'scientific' signifies "the existence of systematic methods of inquiry which when they are brought to bear on a range of facts, enable us to understand them better and to control them more intelligently" (*ibid.*, p. xii). Thus, for Hook there is a "common pattern of inquiry" which is applicable to all fields of study in which claims to truth are made. This applies to the natural and social sciences, technology, history, politics, and ethics. This commitment to the scientific model is an expression of Hook's naturalism, a philosophy that maintains that "all human knowledge is scientific knowledge."[9] This statement, however, is not simply a historical description of what we consider knowledge to be, but a *normative* proposal.

Hook finds scientific method to be embedded in the ways in which we ordinarily act, reason, and function in practical affairs. Hook's approach is not unsympathetic to Thomas Kuhn's, which recognized the impact of culture and society on scientific inquiries. He rejected the epistemological anarchism of Paul Feyerabend, however, because he insisted that there are *some* objective criteria that we use in appraising truth claims, and that knowledge is not simply a matter of taste, caprice, or the slave of the fashions of the day. Hook maintained that "there are working truths on the level of practical living which are everywhere recognized and which everywhere determine the pattern of reasonable conduct" (*ibid.*, p. 173). This conception of prudential rationality is not limited to Western culture. For science simply refines the implicit canons of rationality and intelligi-

[8] *Pragmatism and the Tragic Sense of Life* (New York: Basic, 1974), pp. ix–x.
[9] *The Quest for Being*, p. 216.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

bility involved in the arts and crafts of men and women as they go about the business of living.

Essential to what Hook means by scientific method, common sense, and critical intelligence is the centrality of consequences. We judge hypotheses and theories by their experimental applications, by their observed effects on our conduct and on the material world. The method of inquiry that Hook is talking about thus cannot be justified deductively, nor is it simply assumed as an article of faith. It is tested comparatively by reference to its achievement earned over a long period of human history.

III. CRITIQUE OF SPECULATIVE METAPHYSICS

Hook's pragmatic naturalism can be seen clearly by his criticisms of the metaphysics of Martin Heidegger, Paul Tillich, Jacques Maritain, and others. He attacked the "quest for Being" because "Being" eludes any identifiable confirmation in empirical terms. Hook was not a logical positivist; he rejected its overly narrow conception of philosophy, and especially its indictment of normative philosophy. No doubt influenced by his lifelong colleague Ernest Nagel, he agreed, however, with the positivists' critique of speculative metaphysics. He denied that ontology gives us esoteric knowledge about something or everything that is not found in the sciences. He found the concept of Being, *as such,* to be "non-cognitive," "non-communicative," and a "dispenseful term in discourse" (*ibid.*, p. 155).

Nevertheless, Hook was willing to defend a scientifically developed *Lebensphilosophie.* There are pervasive traits that he thought we could discover in the "cosmic theater of human destiny." Thus, the sciences of human interaction manifest, as Dewey pointed out, "individuality and constant relation," "contingency and need," "movement and arrest," the "stable and precarious." These are not categories of science per se, but reflections on the human condition. Hook's skeptical critique of metaphysics is what led him to reject Marx's form of dialectical materialism. His naturalistic materialism was nonreductive, allowing for a plurality of things and events discovered in various contexts of inquiry.

Unlike the existentialist's *Lebensphilosophie,* Hook had little sense of the irremedial tragic dimensions of human existence. In his *Pragmatism and the Tragic Sense of Life,* he argued that "the universe was open—therefore possibilities were real, the future depended in part upon what human beings did or left undone . . . ideas were potentially plans of action—therefore thinking could and did make a difference to human affairs" (pp. 3–4). Pragmatism, Hook pointed out, is preeminently "a method of *criticism.*" It is, as

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

such, a critique of pretensions and abstractions and bunkum. It seeks to clarify the meaning of ideas by uncovering their consequences in use. By this he was not referring simply to their linguistic uses but to the "behavior of things and people in the concrete situations in which language functions." Pragmatism was "a temper of mind towards the vital options" that we confront. Tragedy in life emerges not because we face death or nonbeing, but because there are conflicts implicit in our moral choices between goods or values, and we are often forced to choose between the lesser of two evils. There is both opportunity and danger in facing an often precarious world; the best method of solving our problems is still, he held, the use of critical intelligence.

IV. HOOK'S NORMATIVE METHODOLOGY

This brings us to the crux of Hook's pragmatic naturalism—the methods by which he proposed we resolve normative problems. Over a long career, Hook concerned himself with innumerable concrete moral issues, such as euthanasia, suicide, civil disobedience, war and peace, law, human rights, quotas, religion, the goals of education, academic freedom, and civil liberties. Unfortunately, he did not always spell out in explicit detail the criteria that he used in testing truth claims. Although throughout his writings he dealt with metamethodological questions, we can best learn about his method, attitude, or temper by seeing him in action, wrestling with concrete practical questions. His methodology is thus implicit in his inquiries.

Hook was committed to naturalistic ethics; that is, he believed that moral and social problems could only be resolved by a deep immersion in the factual content of a situation. In this sense, he differed with so many in the philosophical literature who preferred to deal with logical issues, not their empirical subject matters. We should, said Hook, not be concerned with general naturalistic definitions of value so much as specific values and principles as they emerge in experience. Our decisions could be resolved by using empirical methods similar to those used in the sciences, and by reference to causal and historical knowledge. Here the principle of the continuity of methods comes to the fore. He attempted to relate questions of value to concrete problematic situations. In such questions, we do not start *de nova,* but in the midst of life, drawing upon the fund of knowledge that we have, the values and ends to which we are committed, and the general principles (truthfulness, honesty, etc.) that are taken as conditions of any social life. The situation requires a choice between conflicting norms and values, not the simple deduction from universal rules or first principles.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The process of choice involves deliberations that are related to the attitudes and character of the agent. Hook differed with the emotivists, maintaining that, though attitudes are involved in value judgments, these judgments have a cognitive element. He frequently talked about "ethical knowledge" and "moral truths," which he thought could be discovered by the reflective person within the contexts of critical ethical deliberation. Although we cannot deduce practical judgments from the facts, our choices could best be determined by reference to the relevant facts at hand. A judgment of what should be done is strengthened by the reasons given to support it; its imperative force depends upon the problematic situation, which provides the unifying referent. Where there are disputes about attitudes and values in moral areas, we can only begin to agree if the parties involved are committed to using cognitive methods for adjudicating and negotiating their differences. The genuinely ethical for Hook implies some commitment to rationality.

Hook observed that many have denied that reason is or can be sovereign in human affairs. He conceded that reason often masks underlying passions, but he denied that it is "the function of reason always and only to be the servant of desire." He believed that reason can help us to distinguish between the desired and the desirable; it can reconstruct our desire, hold some desires in check, reinforce others, and still judge between ends that are reasonable and unreasonable. He admits that a reasonable decision in one situation may not be the same in a different situation. That is why each situation requires fresh thinking appropriate to the relevant materials, values, and principles at hand. Hook is an objective relativist in ethics, however, for although our choices are relative to the persons involved in a context, the ends selected need not be subjective but may be objectively modified in the light of critical intelligence. Hook was not an ethical realist, for he did not think that ethical truths had some special status independent of all human cognition, emotions, interests, or needs. Nor was he an intuitionist. In much of this, he agreed with Dewey's general methodology. He differed from him only by the further concretization of ethical problems; he dealt with them not in abstract but in particular terms. The methodological approach that he exemplified was pluralistic in the contexts in which facts, values, principles, and consequences were relevant.

V. PUBLIC AFFAIRS

Hook was skeptical that philosophers *qua* philosophers have any special qualifications in public policy, and he was distrustful of philosophical involvement in public affairs, from Plato to the present, and

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of the follies that some great philosophical minds have displayed. Philosophers, he believed, are far better as *critics* than they are as prophets or seers of social policy. Ordinary men and women deeply involved in public affairs have sometimes displayed greater sagacity than philosophers.

Hook's justification of democracy perhaps best illustrates his approach to such questions. From the standpoint of naturalistic ethics, all claims of facts and values should be considered to be hypotheses, and tested by experience. For Hook democracy as a way of life is justified, if at all, empirically. Hook rejected the skeptical view that democracy is beyond rational justification because it is only the expression of taste and preference. He likewise denied that democracy could be derived from supernatural or metaphysical grounds, or from a doctrine of abstract natural rights. "Why should we treat individuals of unequal talents and endowments as persons who are entitled to relevant consideration and care?"—asked Hook.[10] Because, he replied, democratic methods are more successful in stimulating creative voluntary efforts from members of the community. In enlarging the scope of our experience, democracy affords us the opportunity to appreciate the insights, needs, and aspirations of others. In emphasizing negotiation and compromise, there is apt to be less fear, cruelty, duplicity, more opportunity for a peaceful society, less conflict and strife. In defending free give and take, we maximize the possibilities for discoveries of truth and achieving more reasonable conclusions based on cooperative inquiry. Moreover, democracy will tend in the long run to afford a higher standard of living for a greater number of people, satisfying both economic needs and the demands for cultural freedom.

But why accept *these* values? Hook responds that we should do so not on the basis of any ultimate theoretical principles, or any special metatheory—but because democracy maximizes the widest plurality of values, needs, and interests that a diverse number of citizens share. Now, any one of these values, says Hook, can be disputed: cultural freedom, freedom from cruelty and duplicity, a high standard of living, etc. But each one of these values can be justified in still other contexts of inquiry, without a deduction from ultimate first principles. Hook thus insisted that normative ethical rationality and political wisdom are contextual, appealing to values, rules, and principles, any one of which is open to further empirical justification.

[10] See esp. "Naturalism and Democracy," in Yevant H. Krikoryan, ed., *Naturalism and the Human Spirit* (New York: Columbia, 1944), p. 57.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

VI. SOME CRITICISMS AND REJOINDERS

By way of conclusion, I wish to suggest some criticisms that have been leveled against Hook's pragmatic naturalism and some possible rejoinders.

First objection: Hook presupposed a common pattern of inquiry in both scientific and normative domains, but is it not difficult to elucidate objective criteria that are found in all fields of human interest? Richard Rorty[11] has criticized Hook on this score. Although he applauds Hook and Dewey's defense of a "non-ideological liberalism" at a time when others have abandoned it, he claims that Hook was mistaken in attempting to discover a general method—identified with science—and applicable to all fields of inquiry. Rorty states that, "If one takes the core of pragmatism to be its attempt to replace the notion of true beliefs as representations of 'the nature of things' and instead to think of them as successful rules for action, then it becomes easy to recommend an experimental, fallibilist attitude, but hard to isolate a 'method' that will embody this attitude" (*ibid.*, p. 262).

Rejoinder: Hook is not simply describing how ordinary men and women *invariably* behave in establishing their beliefs. Rather, he is recommending how they *ought* to behave. Pragmatic naturalism thus entails a *normative proposal*. This is consistent with efforts by philosophers historically to set forth an objective method, criteria of rationality, and rules of evidence. Although Rorty may choose to abandon these endeavors, Hook was not prepared to do so. Is it possible to find criteria of objectivity? Surely, our earlier philosophic confidence that there were criteria of objectivity has been weakened, but not entirely. Though clearly, Hook's method reflected an attitude and temper in dealing with questions more than a clearly defined set of principles—he would have disowned any effort to develop these.

The "method of critical intelligence" that Hook defended nevertheless is not alien to what men and women often use in ordinary life. Thus, it is not arbitrary, for it is implicitly embedded in the way we seek to resolve our practical problems. Science is only a more elaborate application of how we think and test our prudential judgments. No doubt, actual human behavior deviates widely from these criteria, not only in religion, ethics, and politics, but even in the sciences. Hook grants this, but he nonetheless recommends that we attempt to use this method as a model and try to extend it wherever possible.

[11] See Richard Rorty, "Pragmatism without Method," in Paul Kurtz, ed., *Sidney Hook: Philosopher of Democracy and Humanism* (Buffalo: Prometheus, 1982).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

But any of the alternative paths offered—faith, authority, mystery, intuition, cultural relativity, subjectivity, etc.—are less effectual, particularly if beliefs are publicly proclaimed as veridical.

Second objection: Hook's criterion of "experimental consequences" is too ambiguous and general. Like other pragmatists, he has never adequately clarified whether he is talking about a theory of meaning or a theory of truth, nor has he clarified sufficiently whether the consequences may be intellectual or whether they must be related to overt behavioral practices.

Rejoinder: This criticism appears to me to have some validity, precisely for the reason that the test of "consequences" is not being used with a univocal or universal signification, but is contextual in meaning and interpretation. Philosophical language by its very nature is so general that it leaves out the specific and concrete. Hook would no doubt reply that by experimental consequences he is talking about the publicly observed results of an inquiry, whether focused on its intellectual or behavioral effects. His interpretation of this is, as I have said, concrete and pluralistic, relative to the varieties of values, principles, and facts in situations.

Third objection: Hook, like Dewey, had too much faith in reason and science, but, in this postmodern age, we have lost confidence in our ability to solve problems by scientific or experimental methods. We need to go beyond the naive optimism of the Golden Age of American Philosophy that "science will save us."

Rejoinder: Hook is much less the optimist than Dewey, having been made acutely aware, perhaps more than others, of the betrayal of the ideals of the Enlightenment by the totalitarian Gulags of the twentieth century. Hook claims to have some sense of the tragic dimensions of experience, though after all is said and done, Hook remained committed to secular humanism, in that he was unwilling to abandon the methodology of scientific reason. He had consistently argued that, however fallible this may be, there are no genuine viable alternatives which we may adopt which are equal in power and promise.

PAUL KURTZ

State University of New York/Buffalo

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

HOOK, DEWEY, AND MARX*

PAUL KURTZ has described Sidney Hook as a "pragmatic naturalist" who, under the influence of Marx and John Dewey, developed a notion of philosophy largely concerned with the relation of theory to practice. In his fallibilist epistemology, in his critique of transcendent metaphysics, and in his writings on public affairs, Hook looked to the mediating role played by norms in the relation between theory and practice. Kurtz has also identified the relation of science to common sense as the key to understanding Hook's philosophy. With all of this I heartily concur.

I shall elaborate upon these contentions from the following unified perspective. For Hook, our understanding of ourselves is more fundamental than our understanding of the physical world. Common sense, as a reflection of how we understand ourselves, is what explains the scientific method. It is not science that explains common sense. Hence, Hook rejected scientism, as well as reductivism, transcendent religion, ontological realism, Platonism, Aristotelianism, idealism, positivism, and analytic philosophy. This is not a rejection of all metaphysics but the embracing of the metaphysics of the Copernican Revolution in philosophy.

For Hook, we are to understand ourselves as responsible social agents, not as detached observers. Practical knowledge is thus more fundamental than theoretical knowledge. This is what Dewey's pragmatism meant for Hook, and this, too, derives from the Copernican Revolution. Unlike animals who deal directly with nature, our actions sustain or modify culture. Our transactions with nature are mediated by culture. There is thus an important break between the natural and the cultural. In this sense, Hook moved beyond a narrow naturalism. Philosophical explication of past social practice leads us to extract from that practice its inherent norms. Hence, any philosophical account we formulate about what we are doing can be used reflexively to decide what to do in the future. In acting out the implications of our "theory," we become more self-conscious and can assume greater responsibility for what we do. By seeing the agent as social, Hook avoided the unfocused methodology that plagues other versions of pragmatism. At the same time, in drawing an analogy in the

* Abstract of a paper to be presented in an APA symposium on the Philosophy of Sidney Hook, December 30, commenting on a paper by Paul Kurtz, this JOURNAL, this issue, 526–534.

0022-362X/90/8710/535–6
© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

way he did between science and common sense, Hook sometimes presented his views as a kind of "theory." The use, by Hook, of the language of "theory" unfortunately obfuscated the sense in which he made fact derivative from value.

The cultural dimension encompasses not only the social world but history as well. Hook saw the agent as a historical being, and this is where Marx played a key role. Dialectic in Marx is the relation between mind and social history. How we think is always in part a reflection of our cultural history. At the same time, thought about the social world leads to action that changes the social world, which, in turn, leads to new thoughts. The relation as expressed in Marx is ambiguous. Many orthodox Marxists see thought as always caused by underlying conditions, thereby making thought essentially derivative.

In Hook's novel version of the dialectic, thought is not simply the product of natural or historical conditions. Neither ideas nor practices are isolable modules. Through interaction, ideas and practices are qualitatively changed so that the history of these interactions is more important than the logic of momentary structure. By recognizing the break between the natural and the cultural, Hook was able to argue that social structure is a reflection of human autonomy. As human beings, we may transcend mere survival in the recognition that we are ultimately self-structuring. This allowed Hook to avoid the pitfalls of objectivism, positivism, and fanaticism. In the end, what Hook left was a moral vision of the evolution of Western Civilization.

NICHOLAS CAPALDI

Queens College, City University of New York

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

SELLARS'S TWO IMAGES OF THE WORLD*

A distinction of fundamental importance to Wilfrid Sellars's philosophy is the one he drew between the manifest and the scientific images of persons in the world.[1] This distinction has puzzled many of his readers, especially those who believe that the task of philosophy is to ascertain the truth about the world rather than contemplate certain images of it. Since so many of Sellars's metaphysical contentions presuppose the distinction, I propose to discuss it here. I want to examine his grounds for insisting upon it and also say something about its philosophical importance.

Sellars believed that the history of philosophy provides a useful means of communication between philosophers, because it offers a well-known body of theory and argumentation to which they can relate one another's claims. This belief is applicable to his distinction between the two images, for anyone who has paid reasonable attention to Hume's philosophy should have some sense of what it is. In both the *Treatise* and the *Enquiry*, Hume[2] drew a distinction between two systems of ideas regarding the world, a popular system and one involved in what he called the modern philosophy; and these systems have at least a family resemblance to Sellars's images. It is true that, unlike Sellars, Hume had a skeptical attitude to both his systems, finding the new one not only indefensible but incomprehensible; but I think his approach to them nevertheless provides a very helpful introduction to the distinction Sellars drew and some of the conclusions he based upon it.

Hume's popular system represents what a contemporary scientific realist would describe as a theoretically naive conception of the world. People accepting this system conceive of the world as a system of observable objects located in a common space and enduring in time. When (as they believe) they observe these objects, they follow a "blind and powerful instinct of nature," Hume said, and suppose "the very images" presented by their senses to be external to them in

* To be presented in an APA symposium on the Philosophy of Wilfrid Sellars, December 29. Jay F. Rosenberg will comment; see this JOURNAL, this issue, 546–7.

[1] Sellars developed this distinction in "Philosophy and the Scientific Image of Man," in *Science, Perception, and Reality* (New York: Humanities, 1963), pp. 1–40. In what follows, I refer to this essay as PSIM.

[2] See *A Treatise of Human Nature*, L. A. Selby-Bigge, ed. (New York: Oxford, 1978), bk I, pt IV, sect. II; and *An Enquiry concerning the Human Understanding*, in *Hume's Enquiries*, L. A. Selby-Bigge, ed. (New York: Oxford, 1966), sect. VII.

0022-362X/90/8710/537–545 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

space.[3] In describing their supposition this way, Hume was using the word 'images' in a referentially transparent or *de re* sense. Expressed a bit differently, his idea was that the items people popularly take to be external objects (for instance, the qualitative expanses they visually apprehend) are, in fact, subjective sensory phenomena, or "images"—and the same is true, *mutatis mutandis,* of the items they might describe as the occurrent smell, sound, taste, or feel of an external object. People who look at a sunset do not suppose (if they use the popular system) that the beautiful purple they see merely corresponds to something in the sky and is the mere effect of a causal "power" that it possesses; they believe it *is* something in the sky, something others can see. Sellars accommodated this sort of belief in his manifest image by taking a pink ice cube as an illustrative example of a "manifest" object. Understood according to his version of the popular system, a pink ice cube is a color solid with tactile and other properties: its pinkness is both occurrent and three-dimensional, a distinctive feature of an external object (PSIM, p. 26).

When he first introduced his manifest image, Sellars described it as the conceptual framework in terms of which human beings first came to be aware of themselves as humans-in-the-world; but he eventually qualified this description, saying that the image is both a historical and a methodological fiction (PSIM, p. 7). His considered opinion seems to have been that human beings (perhaps those in our European tradition) came to conceive of themselves as human beings living in a spatiotemporal, natural world only by developing a conceptual image of themselves and their physical situation which is significantly like the manifest image. Two questions that deserve to be answered here are: "How are we to understand the relevant similarity between actual human conceptions and Sellars's idealized image?" "Why is this idealized image still important, if it is different from a scientific image of human beings and the world?"

A first step in answering these questions is to say that the manifest image is a rational reconstruction of the nontechnical or (as Hume would say) popular system of thought in European culture which antedated the modern philosophical or "corpuscularian" system that Hume found so problematic. Ordinary human beings do not always or perhaps even usually have a unified, determinate conception of the universe and their position in it, but the popular conceptions they do have can be clarified and philosophical sense made of them. The manifest image appears to be Sellars's version of the result of this process. People who accept Hume's popular system are apt to

[3] *Enquiries,* p. 151.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

have many fuzzy or simple-minded views on such subjects as God, their immortal souls, and the location of heaven and hell (one somehow above them, the other somehow below); but Sellars had nothing to say about these topics of popular concern. I expect that he did not include them in his manifest image because he intended the image to be a refinement, as he put it, of actual ways of thinking—one that is disciplined and critical and even, in its way, scientific (PSIM, p. 7). Beliefs about many popular subjects are perhaps too dubious to be included in an image that purports to be "in its way, scientific."

If the manifest image is in a way scientific, how is it distinguished from the scientific image, which is also a fiction or idealization?[4] Sellars was quite specific about this matter. The manifest image is scientific in the sense, he said, that the principal claims involving it can be defended by inductive reasoning of a statistical or correlational kind (and also, no doubt, by appeals to observation and memory). One kind of inductive reasoning is excluded, however: the kind involving "the postulation of imperceptible entities, and principles pertaining to them, to explain the behavior of perceptible things" (PSIM, p. 7). This kind of reasoning is excluded, because Sellars thought it leads to the destruction of the manifest image and the creation of a new system of ideas that, clarified and appropriately reconstructed, is scientific without qualification.

If we concede that there was a transition, in modern European thought, from (in Hume's terms) a popular to a modern philosophical system of ideas and that this transition involved something like the two images of which Sellars spoke, we might want to say that the manifest image is a scientifically obsolete representation of the world, not something that is still worth serious philosophical attention. Sellars would have agreed that the image is scientifically obsolete, but he was convinced that it still deserves our attention. I want to say something about this last matter now. Doing so amounts to answering the second question I asked earlier.

One of Sellars's reasons for insisting on the importance of the manifest image was that it illuminates a persistent tendency in philosophical thinking. Philosophers have always been concerned with the nature of reality; but instead of turning the subject over to the scientists, who are officially dedicated to ascertaining what the world is like and how it is put together, they still try to work out a view of the world in their own terms. In our century, the tendency has been to work out this view by some analytical means—most notably, by analyzing concepts or language. Sellars's remarks on the manifest

[4] Sellars described both images as "idealizations"; see PSIM, p. 5.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

image were allied to this analytical tradition, though his strategy was actually more constructive (or synthetical) than analytical—more Carnapian than late-Wittgensteinian. Instead of reconstructing some portion of scientific language, as Rudolf Carnap was wont to do, Sellars reconstructed a view of the world reflected in, or associated with, common-sense discourse. Not only was ordinary language at the center of attention when he wrote about the manifest image, but he was convinced that the image was an important ingredient in what he called the perennial philosophy.

Sellars described the perennial philosophy as another idealization, one based this time on philosophies that endorse the manifest image as an accurate "large-scale map of reality to which science brings a needle-point of detail and an elaborate technique of map-reading" (PSIM, p. 8). Sellars was himself very critical of the perennial philosophy, but he thought it represented a perennial temptation in philosophy which deserves careful treatment. This temptation leads to trouble in different areas of philosophy, two of which are instructive for appreciating the emphasis Sellars placed on the manifest image.

A basic tenet of any reasonably orthodox empiricist epistemology is that our well-founded beliefs about matters of fact and existence ultimately rest on observation, memory, and inductive (or ampliative) inference. Reasonably orthodox empiricists who accept the perennial philosophy contend that the observations on which our well-founded beliefs about the world are ultimately founded concern manifest objects—that is, colored, fragrant, noisy objects that are frequently resistant to our touch. Statements recording these observations provide the ultimate evidential basis for our beliefs about the unobservable denizens of the physicist's world; suitably corroborated, these beliefs provide the needlepoint of scientific detail which is accepted by the perennial philosophy. The conception of the world resulting from this sort of philosophy is widely accepted by philosophers of common sense, but it bristles with philosophical difficulties. The key difficulty is that the distinctive features of manifest objects are incompatible with a realist interpretation of scientific theory. Aggregates of microparticles may have Lockean powers, but they do not have the occurrent pinkness of Sellars's manifest cube.

Philosophers past and present have attempted to preserve the basic structure of the manifest world by relegating its sensuous content to the domain of mere appearance, but in doing so they open an old-fashioned door to skepticism. If the sensuous qualities apparent in observation are mere appearances, they are subjective phenomena actually belonging to the consciousness of the observer. (The yellow that is apparent to a jaundiced eye belongs to the consciousness, the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

"mind," of the jaundiced observer; it is not a property of the particles making up the paper the observer supposes she is observing.) A skeptical abyss thus opens between the sensing mind and the purely theoretical, nonsensible reality. The observational basis for science is then subjective and nonpublic; and it is hard to see how it can add rational support to conclusions about a fundamentally different, public domain of unobservable objects.[5]

However skeptical they may be about the external world, most philosophers find it impossible to doubt their own existence. But what do they suppose themselves to be? If they are skeptical about the external world, they do not suppose that they are systems of exotic micro-entities; and if they are empiricists, they probably find it hard to suppose that they are Cartesian egos, or spirits. R. M. Chisholm[6] supposes something much more minimal: he supposes, or thinks he knows, that he is an *x* such that. . . . But how can he be so sure that he is a single *x* if (in a preliminary stage of theory construction) he has no idea what the value of his variable is—whether it be spirit or physical body? Bertrand Russell[7] complained that the substance-attribute metaphysics Chisholm is tacitly endorsing here is simply an effect of an old-fashioned subject-predicate logic that is enshrined in our language—a metaphysics that loses its plausibility when we become familiar with the new logic of relations that he and A. N. Whitehead had developed in *Principia Mathematica.* I seriously doubt that the logic of relations has the metaphysical import Russell thought it had in this matter,[8] but I do think that the certainty about one's unitary self belongs to a conception of humans-in-the-world that is intimately related to Sellars's manifest image.

It is easy to see why Sellars insisted that the inductive methods appropriate to the manifest image do not include the method of postulating imperceptible entities to explain the behavior of perceptible things. He insisted on this because he thought that such postulation led to the destruction of the manifest image. To keep the image intact as a workable conceptual framework, one must avoid the introduction of elements that will break it up. This last point is correct, but it does not require the restriction Sellars specified—that is, a ban on using the postulational method in manifest thinking.

[5] Although it is hard to see this, I argue that it can be seen with the right apparatus; cf. my *Knowledge of the External World* (New York: Routledge, in press), chs. 6–8. A contrary claim is ably defended by Benson Mates in *Skeptical Essays* (Chicago: University Press, 1981), pp. 99–153.
[6] *Person and Object* (London: Allen & Unwin, 1976), p. 39.
[7] *The Analysis of Matter* (London: Allen & Unwin, 1927), ch. 23.
[8] I dispute Russell's claim in my *Metaphysics: The Elements* (Minneapolis: Minnesota UP, 1985), p. 80.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

That method is all right in such thinking so long as one does not postulate the wrong entities. The postulation of germs, of small mites on the backs of large mites, or of the insensible parts of objects that Descartes postulated in his *Principles of Philosophy,*[9] does nothing to undermine the reality of Sellars's manifest objects. Imperceptible entities are perfectly allowable, therefore, so long as they are not incompatible with the ultimate reality of manifest objects.

What I have just been saying shows why Sellars thought the perennial philosophy was inherently defective: it tries to unify ideas that, properly clarified, are at odds with one another and belong to different images of the world. It also shows why he thought that the manifest image is important and deserving of critical analysis, even though it is scientifically obsolete. One thing it does not show, however, is what we should actually do with the image once we have appreciated the error in the perennial philosophy. Should we simply cast it aside and adopt an up-to-date scientific alternative?

Sellars's answer was that we cannot adopt such an alternative because one is not yet available. To be sure, we have in quantum theory a highly successful, fine-grained account of physical reality (however strange and paradoxical it may be), but our scientific conception of what might be called mental reality is extremely skimpy and conjectural. Sellars was realistic about this last point; and he neither deplored nor welcomed it. Like most contemporary materialists and functionalists, he was confident that our sensory experiences and our cognitive abilities are owing to the character of our nervous systems and not some fancied immaterial spirit lodged in our bodies like a Cartesian "pilot in a ship"; but he emphasized that our understanding of this matter is still highly rudimentary. To describe our feelings, moods, and sensory experience, and to explain the purposive behavior of ourselves and our neighbors, we have to employ the mentalistic language of everyday life. This language is not only unrelated to the theoretical terms of current neurophysiology; its structure (at least when reasonably clarified) is at odds with the structure of a molecular physical theory. For both descriptive and explanatory reasons pertaining to the mental, we must for the clearly foreseeable future proceed somewhat as Hume did in his *Treatise,* that is, oscillate back and forth between two systems or images. A coherent, all-purpose system is, Sellars insisted, neither available nor even distinctly on the horizon.

Sellars's attitude to an all-purpose system was complicated by the fact that he saw problems with a single system even in a utopian

[9] See pt. four, principle 201, in John Cottingham *et al.*, eds., *Philosophical Writings of Descartes, vol. 1* (New York: Cambridge, 1985), pp. 286–7.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

setting. As he emphasized virtually from the beginning of his career, the phenomena commonly described as mental include items of two significantly disparate classes and raise two distinct mental-physical problems. The phenomena in one of these classes are mental in what he regarded as the proper sense of the term. The mark of such things, he said, is intentionality: properly mental phenomena are conceptual phenomena, things involving logical form and "reference to an object." The other class includes phenomena that are merely sensory, the sort of thing Herbert Feigl called "raw-feels." As ordinarily understood, both sorts of phenomena are attached to unitary subjects, persons; but they raise significantly different problems when one speculates about the relation between them and the entities described in books on neurophysiology.

Philosophers who, in accordance with the temptations of the perennial philosophy, speak of merging what is in effect the manifest with a scientific-theoretical image often contend that popular mentalistic concepts have such minimal content that they can easily fit together with concepts appropriate to the physical sciences. Matters are more complicated than they realize, however. For one thing, we attribute feelings, emotions, intentions, and beliefs to persons, which we take to be single logical subjects. Plausibly interpreted, these mental phenomena belong to the category of *states* (occurrent or dispositional) *of an individual thing*. Since the state of an individual is conceptually different from the state of something (a physical organ) belonging to an individual, or from the state of a system of individuals, a flatfooted mental-physical identity theory is technically objectionable. Sellars's response to this was not to accept an irreducible dualism of mental and physical phenomena but to speculate about what he called successor concepts. If we replace the concept of a unitary individual with the concept of a system of micro-individuals, can we plausibly replace the concept of a individual's sensory state or belief with a concept applicable to a system with some appropriately specified structure? When we ask about making such a replacement, we are, of course, presuming that a satisfactory replacement would not omit anything of theoretical importance.

Sellars found successor concepts for ordinary sensation concepts more problematic than successor concepts for belief, but I am not sure that he was right about this. His opinion on this matter was based on his conviction that ordinary (or manifest) concepts of cognitive states are "nonqualitative": the concept of a thought that triangles are trilateral is, he believed, the concept of an episodic state with a certain conceptual role. Since this latter concept is essentially formal rather than qualitative, there is no evident obstacle to constructing a minimally qualified successor concept which is applicable

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

to a purely physical system (PSIM, pp. 33f.). Abstractly considered, this claim is very plausible; doubts come to mind only when I consider how particular psychological concepts are defined, clarified, or rationally reconstructed. As I reconstruct the concept of believing,[10] for example, actual inferring of a certain sort ought to occur when one acts on a given belief. If physical transitions plausibly corresponding to this inferring do not actually occur when we suppose the belief is being acted upon, we shall have prima facie evidence that a belief explanation for that action is actually misplaced. Although the point is strictly empirical, I think it is conceivable that the concept of a belief explanation and, therefore, the concept of a belief itself might eventually be regarded as scientifically defective and that a technically more advanced theory of human behavior might lack close successors to them.

Sellars found successors to concepts of sensory phenomena particularly problematic because phenomena such as the experience of seeing pink have a continuous qualitative character that is out of place in a gappy system of discrete particles (PSIM, pp. 34–7). Materialists who object that such experiences involve a mere appearance of qualitative continuity simply fan the breeze unless they can make a convincing case for the view that there is nothing continuously qualitative about such appearances. A Cartesian might insist that the relevant appearances are "confused ideas" belonging to a realm actually separate from the physical world, but this alternative is not open to someone who, like Sellars, is a physicalist in the broad sense of the term. Sellars speculated that the problems posed by sensory expanses could ultimately be avoided if (as it actually appears) the ultimate scientific image turns out to be a pure-process rather than a particle theory. He may or not be right about this (I no longer have considered views on the matter), but if the scientific image abjures ultimate particles, it will have moved a long way beyond the system foreshadowed in the modern philosophy that Hume discussed.

Since we do not now have a unified image of ourselves-in-the-world which does justice to the full range of our convictions or considered opinions, Sellars's attitude to the images he identified is something like Hume's attitude to the popular system and his skeptical alternative: we must, in our day-by-day thinking, oscillate back and forth from one to the other. I confess that I commonly do this sort of oscillating. When I contemplate paintings or sunsets, I think of them as occurrently colored and beautiful; I do not, then and

[10] The reconstruction of believing that I favor can be found in my paper, "Action, Inference, Belief, and Intention," in James Tomberlin, ed., *Philosophical Perspectives, vol. 4* (Atascadero: Ridgeview, 1990), pp. 247–271.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

there, picture them as molecular systems with mere sensible powers. Yet I quickly switch back to a more Lockean view when I begin to think about metaphysics. This Lockean view is very incomplete, however, for I am not a physical scientist, experimental or speculative. My scientific image is actually more of a sketch than a map or blueprint. I believe that in basic structure it is closer to the truth than any manifest image, but it is too gappy and fragmentary for my cognitive needs. So, like Hume, I must oscillate back and forth.

When Donald Davidson[11] attacked the very idea of a conceptual scheme, he described his target as something supposedly capable of having alternatives, and he described such supposed alternatives as associated with mutually untranslatable languages. As I have described them, Sellars's two images embody (or employ) distinct conceptual schemes—as do Hume's two systems. Since both images, or both systems, can be described in the English language, distinct conceptual schemes do not have to be associated with mutually untranslatable languages. What they are associated with are distinct systems of belief propositions. When these beliefs are expressed in a common language like English, key words will, of course, be used in different senses—for instance, 'pink' will be used to be mean "occurrently pink" and also (perhaps) to mean "having the power to make normal observers experience pinkness." But there is no general presumption that the idiolects including these different uses of words must be mutually untranslatable—and certainly no presumption that a person using one of these idiolects cannot understand the words of the other. A basic language has simply been developed (systematized and, in one case, extended) in different ways.

Philosophers interested in the subject of alternative conceptual schemes commonly fix their attention on languages differing greatly in structure (such as Hopi and English) and ask themselves whether speakers of the languages view the world differently—and if so, how differently and in exactly what ways. Such philosophers neglect the fact that speakers of the same language have sometimes conceived of the world in fundamentally different ways. I think that these philosophers would do well to ponder Sellars's two images and Hume's two systems: they offer a demystifying approach to a very confusing subject.

BRUCE AUNE

University of Massachusetts/Amherst

[11] "The Very Idea of a Conceptual Scheme," in *Inquiries into Truth and Interpretation* (New York: Oxford, 1974), pp. 183–198.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

RESPONSE TO AUNE*

BRUCE AUNE'S choice of Wilfrid Sellars's distinction between the manifest and the scientific images as a point of entry into the rich and complex Sellarsian philosophy is well-taken, and his comparison of Sellars's two images with Hume's "two systems" is surely useful and revealing. The limitations of the comparison begin to emerge, however, when we note that, unlike Hume, Sellars is unwilling to adopt a skeptical attitude toward the scientific image. It will repay our efforts to discover why this is so.

To frame the issue in another way, Aune asks why, given that it is scientifically obsolete, the manifest image still deserves serious philosophical attention. But the most significant challenges to Sellars's views surely come from the other direction, from advocates of the "perennial philosophy" who endorse the manifest image as real and understand the discourse of science as conceptually dependent upon it, or from radically historicist thinkers who hold that all discourses are created equal and eschew normative epistemology entirely.

The bite of a challenge from this direction, as Sellars recognizes and Aune rightly stresses, is that *persons* are essentially beings who conceive of themselves in terms of the categories of the manifest image, as knowers and doers in a world of causally-interacting spatiotemporal sensible objects. Persons are constituted by a reflexive self-understanding of their own subjectivity, and Humean worries regarding the reality of "external objects" can only arise for such subjective, apperceptive consciousnesses, i.e., ones conceptually equipped with a distinction between "external" and "internal" worlds. It is this *person-constitutive* role which gives Sellars's "manifest image," in contrast to Hume's "popular system of thought," an inescapable philosophical moment.

The scientific image purports to be objective through and through. Science knows only *one* world. Its language is nonindexical and aperspectival, and it recognizes no distinction between "internal" and "external." The true challenge posed by the scientific image is thus that it appears to leave no room for subjectivity and, hence, has no place for *persons*, that is, for *us*. Much of Sellars's philosophical work can be understood as an attempt to show how the

* Abstract of a paper to be presented in an APA symposium on the Philosophy of Wilfrid Sellars, December 30, commenting on a paper by Bruce Aune, this JOURNAL, this issue, 537–545.

0022-362X/90/8710/546–7 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

person-constitutive categories of the manifest image might be reconciled with, or coherently added to, the scientific image to produce a "stereoscopic" or "synoptic" image.[1]

In light of these considerations, the question of why philosophy should take the scientific image seriously in the first place surely becomes increasingly acute. The scientific image, Sellars tells us, emerges from the exercise of a certain kind of postulational theoretical reasoning. The distinction between the two images is thus *au fond* epistemological, and, at least initially, it is not clear why such a distinction should have significant ontological consequences at all, much less give rise to a global philosophical challenge to our own self-conception and self-understanding.

These questions can be answered, if at all, only by bringing into sharper focus just what sort of postulational theoretical reasoning demarcates the scientific from the manifest image, for the application of postulational reasoning per se carries with it no profound philosophical consequences. As Aune insightfully points out, there is nothing problematic about positing, e.g., microbes, and we should recall, too, that the (sensory and intentional) states of persons postulated by Sellars's own mythical genius Jones are central ingredients of the manifest image.

Aune goes on to suggest that what is criterial for the scientific image are postulations of entities that are "incompatible with the ultimate reality of manifest objects." Although this remark is surely right as far as it goes, I cannot help but think that it raises more questions than it answers. Those questions will occupy me in the balance of my response.

JAY F. ROSENBERG

University of North Carolina/Chapel Hill

[1] For a fuller discussion, see my "Fusing the Images," *Journal for General Philosophy of Science*, XXI (1990): 5–23

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

JOHNSONIAN PRIZE IN PHILOSOPHY

The editors are pleased to announce that the Johnsonian Prize, which is co-sponsored by Hackett Publishing Company, and which consists in one thousand dollars as an advance against royalty and guaranteed publication of the prize-winning book, will be offered again in 1991. The prize is designed to encourage the work of unpublished philosophers; authors are eligible who have published no book-length work. Books are selected from among manuscripts sponsored by departments of philosophy.

Previous Johnsonian prize winners have been: 1980, Daniel Bonevac, *Reduction in the Abstract Sciences* (1981); 1981, Igal Kvart, *A Theory of Counterfactuals* (1986); 1982, Moreland Perkins, *Sensing the World* (1983); 1984, Robert J. Dunn, *The Possibility of Weakness of Will* (1986); 1985, Harry van der Linden, *Kantian Ethics and Socialism* (1988); 1986, Clyde L. Hardin, *Color for Philosophers* (1988); 1988, Vann McGee, *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth* (1990). Also, honorable mention in 1981 went to Charles B. Guignon, *Heidegger and the Problem of Knowledge* (1983); and in 1986 to Dionysios Anapolitanos, *Leibniz: Representation, Continuity, and the Spatiotemporal.* Announcement of the eighth prize winner will be made soon in an issue of this JOURNAL.

Authors and departments who wish to submit manuscripts should observe the following guidelines:

1. Any graduate school of philosophy that confers the Ph.D. degree may submit one manuscript per year. It is not required that the work sponsored by a department be written by either a graduate or a faculty member of that department; it may, but need not be a revised version of a departmental dissertation. The closing date for the ninth prize is January 15, 1991.

2. To qualify, the manuscript must be: (a) of book length (at least 40,000 words) and in proper form for publication as a book; (b) accompanied by an abstract of up to 2000 words; (c) written by someone who has not yet published any book-length work; (d) accompanied by two professional evaluations and a letter from the sponsoring department (the editors are depending on the departments to screen out all but the most important and original work); (e) uncommitted to a publisher. Manuscripts will be returned to the authors only if accompanied by appropriate postage.

3. Preference will be given to the work of younger philosophers, especially those who have not yet been granted tenure.

Announcement of the ninth prize winner will be made in a Fall 1991 issue of this JOURNAL; some manuscripts may be given honorable mention. Publication of the prize-winning book will be guaranteed by the Hackett Publishing Company after the author has made what revisions are deemed necessary by the editors. The first seven prize-winning books will be on display with Hackett at the APA Division Meetings in 1990–91.

PLEASE NOTE: The deadline for submission of manuscripts for the Johnsonian Prize for 1991 is *January 15, 1991.*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NOTES AND NEWS

The editors wish to remind our readers that notices of scheduled meetings and symposia are of value only if they appear early enough to make attendance possible. Copy for Notes and News must be planned and processed in this office in time for it to go to the printer eight weeks before publication date, the 1st of each month. Accordingly, an announcement of a meeting to be held, say, in April, should be scheduled for our February issue (or sooner) and should be received in this office no later than December 1, that is, at least four months in advance. The editors appreciate notices that come to us ready for the printer, requiring little editorial work and no retyping, in Journal style (see these columns), with no imperatives, no prices, no telephone numbers, no headings, no unnecessary capitals; and reasonably terse, since space is limited. Also because of space limitations, we cannot guarantee that all announcements sent to us will be published or that those published will appear in the same form (or length) in which they are sent to this office.

Santa Clara University is pleased to announce a conference to be held on April 19–20, 1991. The topic is the Epistemology of Religious Belief. Papers may be submitted for either the conference or the Journal, *LOGOS*, or both. For further information, please write to Elizabeth S. Radcliffe, Philosophy Dept., Santa Clara Univ., Santa Clara, CA 95053.

The Bertrand Russell Society announces a call for papers to be presented at its meeting with the Eastern Division of the American Philosophical Association in December 1991. Papers may be on any aspect of Russell's philosophy. They should have a reading time of about 30 minutes and should be submitted in triplicate with an abstract of not more than 150 words. The name and address of the author and the title of the paper should be on a separate page. Papers should be submitted by April 1, 1991 to David E. Johnson, Chair, Philosopher's Committee, Bertrand Russell Society, Sampson Hall, U.S. Naval Academy, Annapolis, MD 21402-5044.

The North American Division of the Schopenhauer Society announces a call for papers for its seventh annual meeting to be held in conjunction with the Pacific Division Meeting of the American Philosophical Association in March 1991. Papers may be on any aspect of Schopenhauer's philosophy or related themes, and should have a reading time of 20–30 minutes. Papers should be submitted in triplicate to David Cartwright, Philosophy and Religious Studies Dept., Univ. of Wisconsin, Whitewater, WI 53190.

The Stanford Humanities Center will offer up to six external fellowships for 1991–92 intended for postdoctoral scholars and teachers in the humanities, or those in other fields working on related projects, who would be interested in spending the academic year at Stanford. The fellowships are primarily intended to enable fellows to pursue their own research and writing; however, recipients are also expected to devote about one-sixth of their time to

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

teaching or in some other way contributing to intellectual life at Stanford. The deadline for application is December 1. Application materials and further information may be obtained by writing to the Fellowship Program, Stanford Humanities Center, Mariposa House, Stanford Univ., Stanford, CA 94305.

The Department of Philosophy University of California/Riverside, in cooperation with the University of California Humanities Research Institute (Irvine) and the Center for Ideas and Society, is pleased to announce a conference "Skepticism in the History of Philosophy: A Pan-American Dialogue" scheduled for February 15–17, 1991. For more information, please write to the Center for Ideas and Society, Univ. of California, Riverside, CA 92521-0201.

The University Press of Virginia is pleased to announce a new series that will unite scholarly, critical, theoretical, and practical work from such traditionally separate disciplines as anthropology, psychology, history, religion, philosophy, and literature, as well as science, medicine, and law. This interdisciplinary focus within the context of feminist inquiry will invite a reconsideration of every aspect of history and thought. The series editorial board, composed of fifteen feminist scholars representing a variety of disciplines at the University of Virginia, is chaired by Kathleen Balutansky and Alison Booth of the English Dept. Prospective authors are invited to send letters of inquiry and proposals to the Editors, *Feminist Issues: Practice, Politics, Theory,* University Press of Virginia, Box 3608 Univ. Station, Charlottesville, VA 22903.

The Department of Philosophy at Virginia Polytechnic Institute and State University is pleased to announce a conference on the work and thought of Marjorie Grene on the occasion of her eightieth birthday, November 30–December 2. Invited speakers include Alan Donagan, Niles Elredge, Dorothea Frede, and E. Reed. For further information, please write to Joe Pitt, Phil. Dept., Virginia Tech, Blacksburg, VA 24061-0126.

The Department of Philosophy at California State University/Long Beach is pleased to announce the appointments of Julie C. Van Camp and Steven W. Davis as Assistant Professors.

The Department of Philosophy at Rice University is pleased to announce the appointment of Helen E. Longino as Associate Professor of Philosophy and Women's Studies, and Sonja R. Sullivan as Assistant Professor of Philosophy.

The Department of Philosophy at the University of Chicago is pleased to announce that Christine Korsgaard has been promoted to Professor and Michael Forster to Associate Professor, and that Chris Bobonich has been appointed as Assistant Professor.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

AMERICAN PHILOSOPHICAL ASSOCIATION
EASTERN DIVISION
EIGHTY-SEVENTH ANNUAL MEETING

MARRIOTT COPLEY-PLACE HOTEL, BOSTON, MASSACHUSETTS
DECEMBER 27–30, 1990

PROGRAM

THURSDAY, DECEMBER 27

1:00 P.M.–10:00 P.M. REGISTRATION, *Fourth Floor*
1:00 P.M. EXECUTIVE COMMITTEE MEETING, *Rhode Island*

EVENING (8:00–11:00 P.M.)

Group Meetings (see Appendix for Programs)

8:00 CONFERENCE OF PHILOSOPHICAL SOCIETIES, *Brandeis*
8:00 SOCIETY FOR REALIST/ANTIREALIST DISCUSSION, *Suffolk*
8:00 LENSGRINDER'S SOCIETY, *Regis*
8:00 SOCIETY OF CHRISTIAN PHILOSOPHERS, *Vineyard/Falmouth*
8:00 RADICAL PHILOSOPHY ASSOCIATION, *Simmons*
8:00 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Wellesley*

FRIDAY, DECEMBER 28

9:00 A.M.–7:00 P.M. REGISTRATION, *Fourth Floor*
9:00 A.M.–5:00 P.M. BOOK DISPLAYS, *Exhibit Hall*
9:00 A.M.–5:00 P.M. *JOURNAL OF PHILOSOPHY* DISPLAY, *Fourth Floor*

CHILD CARE See the October issue of the *APA Proceedings* for further details.

PLACEMENT INFORMATION

For interviewers, 8:30 A.M.–7:00 P.M., *Maine*
For candidates, 8:30 A.M.–7:00 P.M., *Vermont*

MORNING (9:00–11:00)

Group Meetings (see Appendix for Programs)

9:00 AMERICAN SOCIETY FOR VALUE INQUIRY, *Brandeis*
9:00 SOCIETY FOR REALIST/ANTIREALIST DISCUSSION, *Suffolk*
9:00 SOCIETY FOR THE PHILOSOPHIC STUDY OF GENOCIDE AND THE HOLOCAUST, *Radcliffe*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Friday, December 28, Morning

9:00 INTERNATIONAL ST. THOMAS SOCIETY, *Wellesley*
9:00 SOCIETY FOR THE STUDY OF PROCESS PHILOSOPHIES, *Bentley*
9:00 SOCIETY FOR THE STUDY OF WOMEN PHILOSOPHERS, *Tufts*
9:00 SØREN KIERKEGAARD SOCIETY, *Simmons*
9:00 SOCIETY FOR PHILOSOPHY AND TECHNOLOGY, *Boston College*
9:00 INTERNATIONAL SOCIETY FOR CHINESE PHILOSOPHY, *MIT*
9:00 NORTH AMERICAN NIETZSCHE SOCIETY, *Vineyard/Falmouth*
9:00 RADICAL PHILOSOPHY ASSOCIATION, *Harvard*
9:00 SOCIETY FOR THE PHILOSOPHIC STUDY OF THE CONTEMPORARY VISUAL ARTS, *Northeastern*

LATE MORNING (11:15–1:15)

11:15 AYN RAND SOCIETY, *Suffolk*
11:15 INTERNATIONAL SOCIETY FOR ENVIRONMENTAL ETHICS, *Radcliffe*
11:15 SOCIETY FOR WOMEN IN PHILOSOPHY, *Wellesley*
11:15 SARTRE CIRCLE, *Tufts*
11:15 SOCIETY FOR THE ADVANCEMENT OF AMERICAN PHILOSOPHY *Regis*
11:15 SOCIETY FOR MEDIEVAL AND RENAISSANCE PHILOSOPHY *Falmouth*
11:15 WORLD PHENOMENOLOGY INSTITUTE, *Bentley*
11:15 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Brandeis*

AFTERNOON

2:00–5:00 Session I

A. Symposium: Centennial of William James's *Principles of Psychology, Grand Ballroom E*
Chair: Andrew J. Reck
Speakers: Charlene Haddock Seigfried
Gerald E. Myers
Commentator: H. S. Thayer

B. Symposium: Moral Cognitivism and Realism, *Grand Ballroom F*
Chair: Kurt Baier
Speaker: Stephen Schiffer
Commentators: Nicholas Sturgeon
David B. Wong

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

C. Invited Papers: *Regis*
   1. The Philosophy of Archaeology
      Chair: Merrilee H. Salmon
      Speaker: Alison M. Wylie
      Commentator: James A. Bell
   2. Identity through Time
      Chair: Peter Van Inwagen
      Speaker: Sally Haslanger
      Commentator: George P. Bealer

D. Colloquium: Ethics, *Cape Cod/Hyannis*
   1. *Chair:* Jonathan C. Schonsheck
      Stephen L. Nathanson, "Loyalty, Patriotism, and Impersonal Morality"
      Commentator: Andrew G. Oldenquist
   2. *Chair:* Robert Ehman
      Edward T. Sankowski, "Autonomy and Friendship"
      Commentator: Raymond A. Belliotti
   3. *Chair:* Norman Daniels
      Roger P. Ebertz, "Is Reflective Equilibrium a Coherentist Model?"
      Commentator: Peter Vallentyne

E. Colloquium: Philosophy of Religion, *P Town/Orleans*
   1. *Chair:* George I. Mavrodes
      Laura Garcia, "Functionalism and God"
      Commentator: Terry J. Christlieb
   2. *Chair:* David L. Anderson
      Brian Leftow, "Timelessness and Foreknowledge"
      Commentator: John Martin Fischer
   3. *Chair:* James E. Taylor
      Tomis Kapitan, "Can There Be an Omniscient Agent?"
      Commentator: Richard E. Creel

F. Colloquium: Metaphysics, *Vineyard/Falmouth*
   1. *Chair:* Mark Alfino
      Michael P. Levine, "Historical Antirealism: Boethian Historians Tell their Story"
      Commentator: Paul A. Roth
   2. *Chair:* James J. Fletcher
      Robert Kane, "Free Will, Ultimacy, and Sufficient Reasons"
      Commentator: Donald Sievert

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Friday, December 28, Afternoon, Session I

3. *Chair:* William Ramsey
   James W. Garson, "Cognition without Classical Architecture"
   Commentator: Kent Taylor

G. Colloquium: Plato and Aristotle, *Simmons*

1. *Chair:* Nickolas Pappas
   Jack Montgomery, "Deliberation in the *Crito* (46B-48A)"
   Commentator: Eugenio Benitez
2. *Chair:* Jay Lampert
   James R. Mensch, "Aristotle and the Overcoming of the Subject-object Dichotomy"
   Commentator: David H. Calhoun
3. *Chair:* Thomas O. Buford
   Victor Caston, "Aristotle on Intentionality"
   Commentator: Ronald M. Polansky

H. Special Symposium Arranged by the APA Committee on Computer Use, *Brandeis*
   Topic: Ownership of Computer Software
   Chairs: Deborah G. Johnson and John Ladd
   Speakers: Richard T. DeGeorge, Edwin C. Hettinger, and Richard Stallman

I. Special Symposium Arranged by the APA Committee on Blacks in Philosophy, *Harvard*
   Chair: Thomas Slaughter
   I. Bruce W. Wilshire, "Resistance to Tolerance and Pluralism in World-community"
      Commentator: Claudette Jones
   II. John Pittman, "MacIntyre on Tradition"
      Commentator: Charles Mills
   III. Business Meeting (Lucius Outlaw presiding)

Group Meetings (see Appendix for Programs)

2:00 BERTRAND RUSSELL SOCIETY, *Suffolk*
2:00 SOCIETY FOR BUSINESS ETHICS, *Radcliffe*
2:00 SOCIETY FOR IBERIAN AND LATIN AMERICAN THOUGHT *Wellesley*
2:00 ASSOCIATIONS FOR INFORMAL LOGIC AND CRITICAL THINKING AND PHILOSOPHY OF EDUCATION, *Boston College*
2:00 GANDHI-KING SOCIETY AND CONCERNED PHILOSOPHERS FOR PEACE, *Grand Ballroom C/D*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

LATE AFTERNOON (5:15–7:15)

5:15 INTERNATIONAL BERKELEY SOCIETY, *Brandeis*
5:15 INTERNATIONAL SOCIETY FOR ENVIRONMENTAL ETHICS, *Suffolk*
5:15 HUME SOCIETY, *Radcliffe*
5:15 SOCIETY FOR PHILOSOPHERS AT WORK IN THE WORLD, *Wellesley*
5:15 SOCIETY FOR THE STUDY OF HUSSERL'S PHILOSOPHY, *Bentley*
5:15 SOCIETY FOR ANCIENT GREEK PHILOSOPHY, *Harvard*
5:15 CHARLES S. PEIRCE SOCIETY, *Simmons*
5:15 ASSOCIATION FOR PHILOSOPHY OF THE UNCONSCIOUS *Boston University*
5:15 ASSOCIATION FOR PHILOSOPHY JOURNAL EDITORS *New Hampshire*

EVENING (7:30–10:30)

7:30 LEIBNIZ SOCIETY OF NORTH AMERICA, *Brandeis*
7:30 SOCIETY FOR MEDIEVAL AND RENAISSANCE PHILOSOPHY, *Suffolk*
7:30 NORTH AMERICAN KANT SOCIETY, *Radcliffe*
7:30 JASPERS SOCIETY, *Wellesley*
7:30 SOCIETY FOR ASIAN AND COMPARATIVE PHILOSOPHY, *Bentley*
7:30 PERSONALIST DISCUSSION GROUP, *Tufts*
7:30 CONFERENCE OF PHILOSOPHICAL SOCIETIES, *Regis*
7:30 SOCIETY FOR THE PHILOSOPHIC STUDY OF MARXISM, *MIT*
7:30 SANTAYANA SOCIETY, *Harvard*
7:30 YOUNG PHILOSOPHERS SOCIETY FOR INTERNATIONAL EXCHANGE, *Falmouth*
7:30 ASSOCIATION FOR PHILOSOPHY OF EDUCATION, *Simmons*
7:30 RADICAL PHILOSOPHY ASSOCIATION, *Northeastern*
7:30 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Boston University*

9:00 P.M.–12:00 Midnight RECEPTION, *Grand Ballroom*

SATURDAY, DECEMBER 29

9:00 A.M.–7:00 P.M. REGISTRATION, *Fourth Floor*
9:00 A.M.–5:00 P.M. BOOK DISPLAYS, *Exhibit Hall*
9:00 A.M.–5:00 P.M. *JOURNAL OF PHILOSOPHY* DISPLAY, *Fourth Floor*

PLACEMENT INFORMATION

For interviewers, 8:30 A.M.–7:00 P.M., *Maine*
For candidates, 8:30 A.M.–7:00 P.M., *Vermont*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Saturday, December 29, Morning

MORNING

9:00–11:00 Session II

A. Symposium: The Philosophy of Wilfrid Sellers
Grand Ballroom E

Chair: Robert G. Turnbull
Speaker: Bruce Aune
Commentator: Jay F. Rosenberg

B. Symposium: The Vanishing Subject of Contemporary Discourse: A Pragmatic Response, *Grand Ballroom F*
Chair: Robert S. Sokolowski
Speaker: Vincent M. Colapietro
Commentator: John J. Stuhr

C. Invited Papers, *Cape Cod/Hyannis*
1. Of What Use Is Whitehead's Nontheistic Philosophy?
Chair: Robert Neville
Speaker: Lewis S. Ford
Commentator: Robert S. Corrington
2. Seventeenth-century Women Philosophers and Their Fate in History
Chair: Virginia P. Held
Speaker: Eileen O'Neill
Commentator: Martha B. Bolton

D. Book Discussion: *On Social Fact, P Town/Orleans*
Chair: Lawrence S. Stepelevich
Author: Margaret Gilbert
Commentator: Michael Root

E. Colloquium: Epistemology, *Vineyard/Falmouth*
1. *Chair:* Paul Moser
Murray C. Clarke, "God's Eye Arguments for Epistemic Skepticism"
Commentator: Robert Almeder
2. *Chair:* Thomas M. Olshewsky
Robert French, "The Pragmatic Aspect of Belief"
Commentator: Scott A. Shalkowski

F. Colloquium: Philosophy of Language, *Regis*
1. *Chair:* Mark Richard
Laurence Goldstein, "Pierre, Wittgenstein, and Meaning"
Commentator: Harold I. Brown

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

2. *Chair:* Jaako Hintikka
Brian MacPherson, "Believing in Possibility"
Commentator: John Kern

G. Colloquium: Moral Psychology, *Grand Ballroom A/B*
1. *Chair:* Stephen F. Barker
Alfred R. Mele, "Motivational Ties"
Commentator: Robert L. Frazier
2. *Chair:* John Marshall
David Shier, "Cohen on Human Rationality and Martian Morality"
Commentator: Jonathan Adler

H. Special Symposium Arranged by the APA Committee on the Teaching of Philosophy, *Wellesley*
Topic: Teaching in Ways that Attract Minorities and Women to the Profession
Chair: Mark McPherran
Panel: Howard McGary, Naomi Scheman, Ofelia Schutte, and Laurence Thomas

I. Special Symposium Arranged by the APA Committee on Computer Use, *Northeastern*
Topic: Wittgenstein Archives at the University of Bergen
Chair: Leslie Burkholder
Speakers: Claus Huitfeldt and Ronald Suter

Group Meetings (See Appendix for Programs)

9:00 SOCIETY FOR MEDIEVAL AND RENAISSANCE PHILOSOPHY, *Suffolk*
9:00 SOCIETY FOR THE PHILOSOPHIC STUDY OF GENOCIDE AND THE HOLOCAUST, *Radcliffe*
9:00 WORLD PHENOMENOLOGICAL INSTITUTE, *Bentley*
9:00 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Brandeis*
11:30 A.M.–1:30 P.M. BUSINESS MEETING, *Grand Ballroom*

AFTERNOON

2:15–4:15 Session III

A. Symposium: Realism and Philosophy of Mathematics, *Regis*
Chair: Charles D. Parsons
Speaker: Penelope Maddy
Commentator: Hartry Field

B. Symposium: African Famine: New Economic and Ethical Perspectives, *Vineyard/Falmouth*
Chair: Tom L. Beauchamp
Speaker: George R. Lucas
Commentator: William H. Aiken

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Saturday, December 29, Afternoon, Session III

C. Invited Papers, *P Town/Orleans*
   1. Democracy and Socialism: Philosophical Aporiae
      Chair: Daniel O. Dahlstrom
      Speaker: Frank Cunningham
      Commentator: Elias Baumgarten
   2. The Historical Turn in American Philosophy
      Chair: Douglas M. Kellner
      Speaker: Cornel West
      Commentator: Lorenzo C. Simpson

D. Colloquium: Semantics, *Cape Cod/Hyannis*
   1. *Chair:* William G. Lycan
      Tony Roy, "Linguistic Ersatzism Lives!"
      Commentator: John F. Post
   2. *Chair:* Burton Dreben
      Frank Pecchioni, "Ersatz Completeness in Carnap's *Axiomatik*"
      Commentator: Stephen T. Kuhn

E. Colloquium: Medieval Philosophy, *Northeastern*
   1. *Chair:* James F. Ross
      Scott MacDonald, "Demons, Doxastic Voluntarism, and Aquinas's Account of Faith"
      Commentator: Laura Westra
   2. *Chair:* Paul A. Wagner
      Jeffrey Coombs, "The Possibility of Created Entities in Seventeenth-century Scotism"
      Commentator: Théresè-Anne Druart

F. Colloquium: Applied Ethics, *New Hampshire*
   1. *Chair:* Samuel Freeman
      Michael Gass, "Abortion and Moral Character: A Critique of Smith"
      Commentator: Joram Graf Haber
   2. *Chair:* Timothy Casey
      Bruce Morito, "Holism, Interest-identity, and the Principle of Learned Ignorance"
      Commentator: Craig K. Ihara

G. Special Session Arranged by APA Committee on Teaching Philosophy at Two-Year Colleges, *Suffolk*
   Topic: Teaching Philosophy at the Two-year College: What Graduate School Never Taught Us!
   Participants: To be announced.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

H. Special Symposium Arranged by APA Committees on Philosophy and Medicine and the Status of Women, *Harvard*
Topic: Women, Fetuses, Medicine, and the Law
Chair: Bernard Baumrin
I. Mary Anne Warren, "Sex Selective Abortion"
II. Alan Gewirth, "Fetal Rights"
III. Janet Gallagher, "Women's Rights and Prenatal Decisions"
IV. Mary B. Mahowald, "Women and Fetal Tissue"

Group Meetings (See Appendix for Programs)

2:15 AMERICAN ASSOCIATION FOR THE PHILOSOPHIC STUDY OF SOCIETY, *Brandeis*
2:15 SOCIETY FOR THE PHILOSOPHY OF CREATIVITY, *Radcliffe*
2:15 SOCIETY FOR PHILOSOPHY AND PUBLIC AFFAIRS, *Wellesley*

4:30 PRESIDENTIAL ADDRESS, *Grand Ballroom*
Introduction by Harry Frankfurt
Annette Baier, "A Naturalist View of Persons"

EVENING (6:30–9:30)

Group Meetings (See Appendix for Programs)

6:30 ASSOCIATION FOR PHILOSOPHY AND LIBERATION, *Suffolk*
6:30 INTERNATIONAL SOCIETY FOR NEO-PLATONIC STUDIES, *Radcliffe*
6:30 JASPERS SOCIETY OF NORTH AMERICA, *Wellesley*
6:30 SOCIETY FOR THE PHILOSOPHY OF SEX AND LOVE, *Bentley*
6:30 SOCIETY FOR WOMEN IN PHILOSOPHY, *Simmons*
6:30 SOCIETY FOR THE STUDY OF WOMEN PHILOSOPHERS *Boston University*
6:30 SOCIETY FOR THE PHILOSOPHIC STUDY OF THE CONTEMPORARY VISUAL ARTS, *Boston College*
6:30 CONFERENCE OF PHILOSOPHICAL SOCIETIES, *Regis*
6:30 SOCIETY FOR THE STUDY OF ETHICS AND ANIMALS, *MIT*
6:30 WORLD PHENOMENOLOGICAL INSTITUTE, *Northeastern*
6:30 ASSOCIATION FOR INFORMAL LOGIC AND CRITICAL THINKING *Nantucket*
6:30 RADICAL PHILOSOPHY ASSOCIATION AND APA COMMITTEE ON BLACKS, *Harvard*
6:30 SOCIÉTÉ AMÉRICAINE DE PHILOSOPHIE DE LANGUE FRANÇAISE *Brandeis*

9:00 P.M.–12:00 Midnight RECEPTION, *Grand Ballroom*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

SUNDAY, DECEMBER 30

9:00 A.M.–12:00 Noon REGISTRATION, *Fourth Floor*
9:00 A.M.–2:00 P.M. BOOK DISPLAYS, *Exhibit Hall*
9:00 A.M.–12:00 Noon *JOURNAL OF PHILOSOPHY* DISPLAY, *Fourth Floor*

PLACEMENT INFORMATION

For interviewers, 8:30 A.M.–12:00 Noon, *Maine*
For candidates, 9:00 A.M.–12:00 Noon, *Vermont*

MORNING

9:00–11:00 Session IV

A. Symposium: The Philosophy of Sidney Hook, *Grand Ballroom E*
Chair: Tibor R. Machan
Speaker: Paul Kurtz
Commentator: Nicholas Capaldi

B. Symposium: Cartesian Skepticism and Inference to the Best Explanation, *Grand Ballroom F*
Chair: Ernest Sosa
Speaker: Jonathan M. Vogel
Commentator: Berent Enç

C. Invited Paper: Moral Reasons: A Confucian Perspective, *Regis*
Chair: Ellen Marie Chen
Speaker: Kwong-loi Shun
Commentator: Gregory Trianosky

D. Colloquium: Philosophy of Science, *Cape Cod/Hyannis*
1. *Chair:* Mary B. Mahowald
John R. Hardwig, "Peer Review and Replication Results: Obsolescent Safeguards in Science?"
Commentator: Joan C. Callahan
2. *Chair:* Ernan McMullin
Paul C. L. Tang and Erik A. Hildebrand, "Quantum Mechanics, Value Judgments, and Scientific Realism"
Commentator: Robert Schwartz

E. Colloquium: Modern Philosophy, *P Town/Orleans*
1. *Chair:* Craig Walton
Kent B. Linville, "Belief, Doubt, and Former Opinions in Descartes's First Meditation"
Commentator: Robert Hanna

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

2. *Chair:* Steven G. Crowell
   Richard McCarty, "Internalism, Externalism, and Kant's Practical Synthesis"
   Commentator: Jane Kneller

F. Colloquium: American Philosophy, *Vineyard/Falmouth*
1. *Chair:* Roberta Kevelson
   Stuart E. Rosenbaum, "Truth and Evaluation"
   Commentator: Christopher B. Kulp
2. *Chair:* Thelma Z. Lavine
   Gregory Pappas, "William James and the Logic of Faith" (The William James Prize Essay)
   Commentator: Eugene Fontinell

G. Colloquium: Time and Hermeneutics, *Grand Ballroom A/B*
1. *Chair:* George L. Kline
   Jerome A. Miller, "Time as Hemorrhage"
   Commentator: Charles Sherover
2. *Chair:* John D. Caputo
   Shaun Gallagher, "The Place of *Phronesis* in Postmodern Hermeneutics"
   Commentator: Bina Gupta

Group Meetings (see Appendix for Programs)

9:00 SOCIETY FOR WOMEN IN PHILOSOPHY, *Wellesley*
9:00 AMERICAN SOCIETY FOR VALUE INQUIRY, *Brandeis*
9:00 SARTRE CIRCLE, *Suffolk*
11:15 A.M.–12:15 P.M. BUSINESS MEETING (Second Session, if needed)
Grand Ballroom

AFTERNOON

1:30–4:30 Session V

A. Symposium: John Locke after Three Hundred Years
Grand Ballroom E

Chair: Margaret Atherton
Speakers: John W. Yolton
James Tully
Commentator: Vere Chappell

B. Symposium: Gender, Race, Ethnicity: Sources of Identity
Grand Ballroom F

Chair: Bernard R. Boxill
Speakers: Anthony Appiah
Maria C. Lugones
Commentator: Thomas E. Wartenberg

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Sunday, December 30, Afternoon, Session V

C. Book Discussion: *The Moral Collapse of the University* *Vineyard/Falmouth*

Chair: Thomas K. Hearn
Author: Bruce Wilshire
Commentators: Myles Brand
Michael K. Hooker

D. Colloquium: Ethics, II, *Grand Ballroom A/B*

1. *Chair:* William Gerber
Bruce D. Weinstein, "The Possibility of Ethical Expertise"
Commentator: Lisa Parker
2. *Chair:* Michael R. DePaul
Richard K. Greenstein, "Hard Cases and Moral Dilemmas: An Essay concerning Legal and Moral Reasoning"
Commentator: Judith W. DeCew
3. *Chair:* Evelyn M. Barker
Linda L. McAlister, "Joy in Ambiguity: A Virtue for Our Time"
Commentator: Marilyn Friedman

E. Colloquium: Social and Political Philosophy *Grand Ballroom C/D*

1. *Chair:* Janet Coy
Sterling Harwood, "Affirmative Action is Justified: A Reply to Newton"
Commentator: Lisa Newton
2. *Chair:* James S. Gouinlock
Robert N. Van Wyk, "Sex Education, the Family, the State, and Political Theory"
Commentator: Alan Soble
3. *Chair:* Robert Paul Churchill
John H. Bogart, "Crime, Authority, and the Moral Standing of the State"
Commentator: Win-chiat Lee

F. Colloquium: Merleau-Ponty, Sartre, and Foucault, *Regis*

1. *Chair:* Rod Bertolet
Lawrence Hass, "The Antinomy of Perception: Merleau-Ponty and Causal Representation Theory"
Commentator: Osborne P. Wiggins
2. *Chair:* Thomas R. Flynn
Robert W. Switzer, "Metaphysical Shame: Reflections on Sartre's Encounter with the Other"
Commentator: Thomas C. Anderson

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

3. *Chair:* Charles Scott
Mario Moussa, "Foucault and the Problem of Agency; or, Toward a Practical Philosophy"
Commentator: Stephen Watson

G. Special Session Arranged by the APA Committee on International Cooperation, *Harvard*
Topic: Philosophy in Yugoslavia Today
Chair: James P. Scanlan
Speaker: Matjaž Potrč, "Philosophy of Psychology in Slovenia and Yugoslavia"

H. Special Symposium Arranged by the APA Committee on the Status of Women, *Simmons*
Topic: Doing Philosophy as a Feminist
Chair: Louise Antony
Speakers: Sally Haslanger, Virginia Held, Mary Mothersill, and Naomi Scheman

1:30 EXECUTIVE COMMITTEE MEETING, *Rhode Island*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

PROGRAM PARTICIPANTS
EIGHTY-SEVENTH ANNUAL MEETING

ADLER, Jonathan, City University of New York/Brooklyn College II-G
AIKEN, William H., Chatham College, Pennsylvania III-B
ALFINO, Mark, Gonzaga University I-F
ALMEDER, Robert, Georgia State University II-E
ANDERSON, David L., Illinois State University I-E
ANDERSON, Thomas C., Marquette University V-F
ANTONY, Louise, North Carolina State University V-H
APPIAH, Anthony, Cornell University V-B
ATHERTON, Margaret, University of Wisconsin/Milwaukee V-A
AUNE, Bruce, University of Massachusetts/Amherst II-A
BAIER, Annette, University of Pittsburgh Presidential Address
BAIER, Kurt, University of Pittsburgh I-B
BARKER, Evelyn M., University of Maryland/Baltimore City V-D
BARKER, Stephen F., Johns Hopkins University II-G
BAUMGARTEN, Elias, University of Michigan/Dearborn III-C
BAUMRIN, Bernard, City University of New York/Lehman College III-H
BEALER, George P., Reed College, Oregon I-C
BEAUCHAMP, Tom L., Georgetown University III-B
BELL, James A., University of South Florida I-C
BELLIOTTI, Raymond A., State University of New York/Fredonia I-D
BENITEZ, Eugenio, The Catholic University of America I-G
BERTOLET, Rod, Purdue University V-F
BOGART, John H., University of Southern California V-E
BOLTON, Martha B., Rutgers University/New Brunswick II-C
BOXILL, Bernard R., University of North Carolina/Chapel Hill V-B
BRAND, Myles, University of Oregon V-C
BROWN, Harold I., Northern Illinois University II-F
BUFORD, Thomas O., Furman University I-G
BURKHOLDER, Leslie, Carnegie Mellon University II-I
CALHOUN, David H., Colby College, Maine I-G
CALLAHAN, Joan C., University of Kentucky IV-D
CAPALDI, Nicholas, City University of New York/Queens College IV-A
CAPUTO, John D., Villanova University IV-G
CASEY, Timothy, University of Scranton III-F
CASTON, Victor, University of Texas/Austin I-G
CHAPPELL, Vere, University of Massachusetts/Amherst V-A
CHEN, Ellen M., St. John's University IV-C
CHRISTLIEB, Terry J., Tarkio College, Missouri I-E
CHURCHILL, Robert Paul, George Washington University V-E
CLARKE, Murray C., Concordia University II-E
COLAPIETRO, Vincent M., Fordham University II-B
COOMBS, Jeffrey, Austin, Texas III-E
CORRINGTON, Robert S., Drew University II-C
COY, Janet, Health Science Center, State University of New York/Syracuse V-E

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

CREEL, Richard E., Ithaca College, New York I-E
CROWELL, Steven G., Rice University IV-E
CUNNINGHAM, Frank, University of Toronto, Canada III-C
DAHLSTROM, Daniel O., The Catholic University of America III-C
DANIELS, Norman, Tufts University I-D
DECEW, Judith W., Clark University V-D
DeGEORGE, Richard T., University of Kansas I-H
DEPAUL, Michael R., University of Notre Dame V-D
DREBEN, Burton, Harvard University III-D
DRUART, Thérese-Anne, The Catholic University of America III-E
EBERTZ, Roger P., West Virginia University I-D
EHMAN, Robert, Vanderbilt University I-D
ENÇ, Berent, University of Wisconsin/Madison IV-B
FIELD, Hartry, University of Southern California III-A
FISCHER, John Martin, University of California/Riverside I-E
FLETCHER, James J., George Mason University I-F
FLYNN, Thomas R., Emory University V-F
FONTINELL, Eugene, City University of New York/Queens College IV-F
FORD, Lewis S., Old Dominion University II-C
FRAZIER, Robert L., Virginia Commonwealth University II-G
FREEMAN, Samuel, University of Pennsylvania III-F
FRENCH, Robert, Ohio Wesleyan University II-E
FRIEDMAN, Marilyn, Purdue University V-D
GALLAGHER, Janet, New York City III-H
GALLAGHER, Shaun, Canisius College, New York IV-G
GARCIA, Laura, Georgetown University I-E
GARSON, James W., University of Houston I-F
GASS, Michael, Kent, Ohio III-F
GERBER, William, Washington, DC V-D
GEWIRTH, Alan, University of Chicago III-H
GILBERT, Margaret, University of Connecticut/Storrs II-D
GOLDSTEIN, Laurence, Hong Kong University II-F
GOUINLOCK, James S., Emory University V-E
GREENSTEIN, Richard K., Temple University V-D
GUPTA, Bina, University of Missouri/Columbia IV-G
HABER, Joram Graf, Bergen City College, New Jersey III-F
HANNA, Robert, University of Colorado/Boulder IV-E
HARDWIG, John R., East Tennessee State University IV-D
HARWOOD, Sterling, San Jose State University V-E
HASLANGER, Sally, University of Pennsylvania I-C, V-H
HASS, Lawrence, University of Illinois/Urbana-Champaign V-F
HEARN, Thomas K., Wake Forest University V-C
HELD, Virginia P., City University of New York/Hunter College II-C, V-H
HETTINGER, Edwin C., College of Charleston, South Carolina I-H
HILDEBRAND, Erik A., California State University/Long Beach IV-D
HINTIKKA, Jaako, Boston University II-F
HOOKER, Michael K., University of Maryland/Baltimore County V-C

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

HUITFELDT, Claus, University of Bergen, Norway II-I
IHARA, Craig K., California State University/Fullerton III-F
JOHNSON, Deborah G., Rensselaer Polytechnic Institute, New York I-H
JONES, Claudette, University of the District of Columbia I-I
KANE, Robert, University of Texas/Austin I-F
KAPITAN, Tomis, East Carolina University I-E
KELLNER, Douglas M., University of Texas/Austin III-C
KERN, John, State University of New York/Buffalo II-F
KEVELSON, Roberta, Pennsylvania State University IV-F
KLINE, George L., Bryn Mawr College, Pennsylvania IV-G
KNELLER, Jane, Colorado State University/Fort Collins IV-E
KUHN, Stephen T., Georgetown University III-D
KULP, Christopher B., Santa Clara University IV-F
KURTZ, Paul, State University of New York/Buffalo IV-A
LADD, John, Brown University I-H
LAMPERT, Jay, Howard University I-G
LAVINE, Thelma Z., George Mason University IV-F
LEE, Win-chiat, Wake Forest University V-E
LEFTOW, Brian, Fordham University I-E
LEVINE, Michael P., University of Western Australia I-F
LINVILLE, Kent B., Oxford College, Georgia IV-E
LUCAS, George R., Clemson University III-B
LUGONES, Maria C., Carleton College, Minnesota V-B
LYCAN, William G., University of North Carolina/Chapel Hill III-D
MacDONALD, Scott, University of Iowa III-E
MACHAN, Tibor R., Auburn University IV-A
MacPHERSON, Brian, Windsor, Ontario, Canada II-F
MADDY, Penelope, University of California/Irvine III-A
MAHOWALD, Mary B., University of Chicago III-H, IV-D
MARSHALL, John, University of Virginia II-G
MAVRODES, George I., University of Michigan I-E
McALISTER, Linda L., University of South Florida V-D
McCARTY, Richard, East Carolina University IV-E
McGARY, Howard, Rutgers University/New Brunswick II-H
McMULLIN, Ernan, University of Notre Dame IV-D
McPHERRAN, Mark, University of Maine/Farmington II-H
MELE, Alfred R., Davidson College, North Carolina II-G
MENSCH, James R., Nova Scotia I-G
MILLER, Jerome A., Salisbury State College, Maryland IV-G
MILLS, Charles, University of Illinois/Chicago I-I
MONTGOMERY, Jack, Saint Xavier College, Illinois I-G
MORITO, Bruce, University of Victoria, Canada III-F
MOSER, Paul, Loyola University of Chicago II-E
MOTHERSILL, Mary, Columbia University/Barnard College V-H
MOUSSA, Mario, Worcester Polytechnic Institute V-F
MYERS, Gerald E., City University of New York/Queens College I-A
NATHANSON, Stephen L., Northeastern University I-D

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NEVILLE, Robert, Boston University II-C
NEWTON, Lisa, Fairfield University V-E
O'NEILL, Eileen, University of Pennsylvania II-C
OLDENQUIST, Andrew G., Ohio State University I-D
OLSHEWSKY, Thomas M., University of Kentucky II-E
PAPPAS, Gregory, University of Texas/Austin IV-F
PAPPAS, Nickolas, Hollins College, Virginia I-G
PARKER, Lisa, University of Pittsburgh V-D
PARSONS, Charles D., Harvard University III-A
PECCHIONI, Frank, Jefferson Community College, Kentucky III-D
PITTMAN, John, City University of New York/John Jay College I-I
POLANSKY, Ronald M., Duquesne University I-G
POST, John F., Vanderbilt University III-D
POTRČ, MATJAŽ, Yugoslavia V-G
RAMSEY, William, Notre Dame University I-F
RECK, Andrew J., Tulane University I-A
RICHARD, Mark, Tufts University II-F
ROOT, Michael, University of Minnesota II-D
ROSENBAUM, Stuart E., Baylor University IV-F
ROSENBERG, Jay F., University of North Carolina II-A
ROSS, James F., University of Pennsylvania III-E
ROTH, Paul A., University of Missouri/St. Louis I-F
ROY, Tony, University of California/Davis III-D
SALMON, Merrilee H., University of Pittsburgh I-C
SANKOWSKI, Edward T., University of Oklahoma I-D
SCANLAN, James P., Ohio State University V-G
SCHEMAN, Naomi, University of Minnesota/Minneapolis II-H, V-H
SCHIFFER, Stephen, City University of New York/Graduate School I-B
SCHONSHECK, Jonathan C., Le Mayne College, New York I-D
SCHUTTE, Ofelia, University of Florida II-H
SCHWARTZ, Robert, University of Wisconsin/Milwaukee IV-D
SCOTT, Charles, Vanderbilt University V-F
SEIGFRIED, Charlene Haddock, Purdue University I-A
SHALKOWSKI, Scott A., University of Oklahoma II-E
SHEROVER, Charles, City University of New York/Hunter College IV-G
SHIER, David, Wayne State University II-G
SHUN, Kwong-loi, University of California/Berkeley IV-C
SIEVERT, Donald, University of Missouri/Columbia I-F
SIMPSON, Lorenzo C., University of Richmond III-C
SLAUGHTER, Thomas, Bentley College, Massachusetts I-I
SOBLE, Alan, University of New Orleans V-E
SOKOLOWSKI, Robert S., The Catholic University of America II-B
SOSA, Ernest, Brown University IV-B
STALLMAN, Richard, Free Software Foundation I-H
STEPELEVICH, Lawrence S., Villanova University II-D
STUHR, John J., University of Oregon II-B
STURGEON, Nicholas, Cornell University I-B

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

SUTER, Ronald, Michigan State University II-I
SWITZER, Robert W., Queen's University, Canada V-F
TANG, Paul C. L., California State University/Long Beach IV-D
TAYLOR, James E., Bowling Green State University I-E
TAYLOR, Kent, University of Maryland I-F
THAYER, H. S., City University of New York/City College I-A
THOMAS, Laurence, Syracuse University II-H
TRIANOSKY, Gregory, University of Michigan/Flint IV-C
TULLY, James, McGill University, Canada V-A
TURNBULL, Robert G., Ohio State University II-A
VALLENTYNE, Peter, Virginia Commonwealth University I-D
VAN INWAGEN, Peter, Syracuse University I-C
VAN WYK, Robert N., University of Pittsburgh V-E
VOGEL, Jonathan M., Amherst College/Massachusetts IV-B
WAGNER, Paul A., University of Houston/Clear Lake III-E
WALTON, Craig, University of Nevada IV-E
WARREN, Mary Anne, San Francisco State College, California III-H
WARTENBERG, Thomas E., Mount Holyoke College, Massachusetts V-B
WATSON, Stephen, University of Notre Dame V-F
WEINSTEIN, Bruce D., West Virginia University V-D
WEST, Cornel, Princeton University III-C
WESTRA, Laura, Auburn University III-E
WIGGINS, Osborne P., University of Louisville V-F
WILSHIRE, Bruce W., Rutgers University/New Brunswick I-I, V-C
WONG, David B., Brandeis University I-B
WYLIE, Alison M., University of Western Ontario, Canada I-C
YOLTON, John W., Rutgers University/New Brunswick V-A

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

APPENDIX: PROGRAMS OF GROUP MEETINGS

THURSDAY, DECEMBER 27

I. EVENING (8:00–11:00)

1. 8:00 CONFERENCE OF PHILOSOPHICAL SOCIETIES, *Brandeis*
   Chair: John A. Loughney
   Annual Board Meeting

2. 8:00 SOCIETY FOR REALIST/ANTIREALIST DISCUSSION, *Suffolk*
   Topic: Reflections on C. S. Peirce
   I. Peter Ochs, "Joy in Work; Joy in Creativity"
   II. Mary Carmen Rose, "The Love of Truth"

3. 8:00 LENSGRINDERS'S SOCIETY, *Regis*
   Reception for Job Seekers

4. 8:00 SOCIETY OF CHRISTIAN PHILOSOPHERS, *Vineyard/Falmouth*
   Thomas Tracy, "Action: Created and Divine"
   Commentator: Derk Pereboom

5. 8:00 RADICAL PHILOSOPHY ASSOCIATION, *Simmons*
   Topic: Nationalism, Internationalism, and Self-determination
   Chair: Richard Schmitt
   Omar Dahbour, "Critique of National Self-determinism"
   Commentator: Phil Gasper

6. 8:00 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Wellesley*
   Topic: Corrective Justice and Abstract Right
   Chair: Richard D. Winfield
   Peter Benson, "The Independence of Corrective Justice and its Priority: On The Significance of Abstract Right"
   Commentator: Joshua Cohen

FRIDAY, DECEMBER 28

II. MORNING (9:00–11:00)

1. 9:00 AMERICAN SOCIETY FOR VALUE INQUIRY, *Brandeis*
   Topic: Presidential Address
   Chair: John Abbarno
   I. Thomas Magnell, "Evaluations as Assessments"
   II. Business Meeting

2. 9:00 SOCIETY FOR REALIST/ANTIREALIST DISCUSSION, *Suffolk*
   Topic: Present Realist/Antirealist Issues
   I. Kendall D'Andrade, "A Republic without Realism"
   II. Raymond J. Langley, "The Synthetic Dialectic between Illusion and Reality in Christopher Caudwell"

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Friday, December 28, Morning, II

3. 9:00 SOCIETY FOR THE PHILOSOPHIC STUDY OF GENOCIDE AND THE HOLOCAUST, *Radcliffe*
Chair: Alan Rosenberg
I. James Bernauer, "On the Meaning of Nazi Ethics for Philosophy"
II. Peter Haas, "Fashioning An Ethical Discourse: The Case of Nazi Germany"
III. Andre Mineau, "Ethics in Auschwitz: Haas and Nazi Morality"

4. 9:00 INTERNATIONAL ST. THOMAS SOCIETY, *Wellesley*
Chair: Mary Rose Barral
John Nota, "The Importance and Relevance of Knowledge by Connaturality"

5. 9:00 SOCIETY FOR THE STUDY OF PROCESS PHILOSOPHIES, *Bentley*
Chair: Joseph Grange
Topic: Robert Neville's *Recovery of the Measure*
George Allan, "Measuring the Recovery"
Commentator: Robert Neville

6. 9:00 SOCIETY FOR THE STUDY OF WOMEN PHILOSOPHERS, *Tufts*

7. 9:00 SØREN KIERKEGAARD SOCIETY, *Simmons*
Chair: John Donnelly
I. Bernard Zelechow, "Kierkegaard, the Aesthetic, and Mozart's *Don Juan*"
Commentator: Ronald Hall
II. George Pattison and Badwell Ash, "Subjectivity: Point of View, Language Game, or Form of Life?"
Commentator: Charles Creegan

8. 9:00 SOCIETY FOR PHILOSOPHY AND TECHNOLOGY, *Boston College*
Topic: Technology and Culture
Chair: Stanley R. Carpenter
Panel: Paul T. Durbin, David Lovekin, and Marx Wartofsky

9. 9:00 INTERNATIONAL SOCIETY FOR CHINESE PHILOSOPHY, *MIT*
Chair and Commentator: Kwang-Sae Lee
I. Sidney Axinn, "The Philosophy of History in Confucianism and in Kant"
II. Xiaorong Li, "The Paradox of Value"
III. Jeffrey Wattles, "From Self-examination to Moral Spontaneity with the Confucian Golden Rule"

10. 9:00 NORTH AMERICAN NIETZSCHE SOCIETY, *Vineyard/Falmouth*
Topic: Nietzsche and Emerson
Chair: Richard Fleming
Speakers: James Conant, Tim Gould, Karen Hanson, and Stanley Cavell
Annual Business Meeting

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

11. 9:00 RADICAL PHILOSOPHY ASSOCIATION, *Harvard*
Topic: The Relevance of Philosophy to Political Activism
Chair: Roger Gottlieb
Speakers: Kathryn Pyne Addelson and Steven Jay Gold

12. 9:00 SOCIETY FOR THE PHILOSOPHIC STUDY OF THE CONTEMPORARY VISUAL ARTS, *Northeastern*
Chair: John A. Loughney
I. Allan Casebier, "A Phenomenological Analysis of Cinematic Sound"
II. Sander Lee, "Alfred Hitchcock: Misogynist or Feminist?"
III. Douglas Kellner, "Reading Film Politically"

III. LATE MORNING (11:15–1:15)

1. 11:15 AYN RAND SOCIETY, *Suffolk*
Topic: Ayn Rand on Value and Obligation
Chair: James G. Lennox
Speakers: Douglas Rasmussen and Allan Gotthelf

2. 11:15 INTERNATIONAL SOCIETY FOR ENVIRONMENTAL ETHICS, *Radcliffe*
Topic: Environmental Ethics: Current Trends, Future Prospects
Chair: Holmes Rolston
Panel: Andrew Brennan, Sara Ebenreck, Bryan Norton, Holmes Rolston, and Mark Sagoff

3. 11:15 SOCIETY FOR WOMEN IN PHILOSOPHY, *Wellesley*
Topic: Feminism and Moral Philosophy
Chair: Jana Sawicki
I. Kathryn Pyne Addelson, "What is a Feminist Moral Problem: A Question of Moral Truth and Moral Action"
II. Marilyn Friedman, "They Lived Happily Ever after: Sommers on Women and Marriage"
Commentator: Cristina H. Sommers

4. 11:15 SARTRE CIRCLE, *Tufts*
Chair: Phyllis S. Morris
I. Eleanore Holveck, "Sartre's Treatment of Women: The Personal, the Political, and the Philosophical"
Commentator: Eleanor Kuykendall
II. Peter McInerney, "Being Responsible for Unexamined Ends"
Commentator: David Detmer

5. 11:15 SOCIETY FOR THE ADVANCEMENT OF AMERICAN PHILOSOPHY *Regis*
Topic: Cornel West's *The American Evasion of Philosophy*
Chair: William J. Gavin
I. Konstantin Kolenda, "The (Cornel) West-ward Vision for American Philosophy

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Friday, December 28, Late Morning, III

II. Robert S. Corrington, "The Emancipation of American Philosophy"
Respondent: Cornel West

6. 11:15 SOCIETY FOR MEDIEVAL AND RENAISSANCE PHILOSOPHY, *Falmouth*
Executive Committee Meeting

7. 11:15 WORLD PHENOMENOLOGY INSTITUTE, *Bentley*
Chair: Christopher Braider
I. Thomas Ryba, "The Category of the Sacred: Four Views"
II. Robert Williams, "Some Reflections about the Sacred"
III. Mary Rose Barral, "The New Approach to the Sacred in Recent Phenomenology"

8. 11:15 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Brandeis*
Topic: The Twilight of Postmodernism
Chair: Chip Sills
I. William Maker, "(Postmodern) Tales from the Crypt: The Night of Zombie Philosophers"
II. Carl Rapp, "After Philosophy?: The Crisis of Reason in Contemporary Thought"

IV. AFTERNOON (2:00–5:00)

1. 2:00 BERTRAND RUSSELL SOCIETY, *Suffolk*
Chair: Marvin Kohl
I. Jan Dejnozka, "Russell's Fifteen Private-language Arguments"
Commentator: Fred Guy
II. Trip McCrossin, "Bundle Theory: Its Scope and Limits"
Commentator: Albert Casullo

2. 2:00 SOCIETY FOR BUSINESS ETHICS, *Radcliffe*
Chair: Jennifer Moore

3. 2:00 SOCIETY FOR IBERIAN AND LATIN AMERICAN THOUGHT, *Wellesley*
Chair: Jorge Gracia
I. Vicente Medina, "Latin American Philosophy: Fact, Fiction, or Utopia"
II. Amy Oliver, "Contemporary Latin American Philosophy: The Polemics of Philosophical Category"
III. Anne Ashbaugh, "Women in Love, Women in Pain: The Female Perspective on Love in the Novels of Monserrat Roig"
Commentator: Ofelia Schutte
IV. Business Meeting

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

4. 2:00 ASSOCIATIONS FOR INFORMAL LOGIC AND CRITICAL THINKING AND PHILOSOPHY OF EDUCATION, *Boston College*
Topic: Israel Scheffler's Contributions to Philosophy and Education: A Critical Appreciation
Chair: Nelson Goodman
I. Barbara Arnstine and Donald Arnstine, "The Contribution of Reasons to Reasonable Practice: Scheffler's Philosophy of Education"
II. Catherine Z. Elgin, "Scheffler's Symbols"
III. Karen Hanson, "Reconstruction in Pragmatism"
Respondent: Israel Scheffler
IV. Business Meeting (4:00 P.M.)

5. 2:00 GANDHI-KING SOCIETY AND CONCERNED PHILOSOPHERS FOR PEACE, *Grand Ballroom C/D*
Topic: Pragmatic versus Principled Nonviolence
Chair: Richard Werner
Speakers: Gene Sharp and Sissela Bok

V. LATE AFTERNOON (5:15–7:15)

1. 5:15 INTERNATIONAL BERKELEY SOCIETY, *Brandeis*
Chair: Kenneth P. Winkler
Phillip D. Cummins, "Berkeley's Manifest Qualities Thesis" (1990 Colin and Ailsa Turbayne Berkeley Prize Essay)
Commentator: James Van Cleve

2. 5:15 INTERNATIONAL SOCIETY FOR ENVIRONMENTAL ETHICS, *Suffolk*
Topic: Issues in Environmental Ethics
Chair: Eric Katz
Commentators: Andrew Brennan and William Aiken

3. 5:15 HUME SOCIETY, *Radcliffe*
Chair: Eric Steinberg
Speaker: Saul Traiger

4. 5:15 SOCIETY FOR PHILOSOPHERS AT WORK IN THE WORLD, *Wellesley*
I. Mary Mulhern, "Conceptual Tools for Forging New Corporate Identities"
II. Nance Cunningham and J. David Newell, "Labeling Patients: The Ethical Consequences of Violating Identity"

5. 5:15 SOCIETY FOR THE STUDY OF HUSSERL'S PHILOSOPHY, *Bentley*
Topic: Husserl on Representation
Chair: Allan Casebier
Panel: David Smith, Dallas Willard, J. N. Mohanty, and Peter McCormick

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Friday, December 28, Late Afternoon, V

6. 5:15 SOCIETY FOR ANCIENT GREEK PHILOSOPHY, *Harvard*
Topic: Plato's *Timaeus*
Chair: Anthony Preus
I. Thomas M. Robinson, "The Theodicy of the *Timaeus*"
II. Dorothea Frede, "Common Concepts in Plato's *Timaeus*"

7. 5:15 CHARLES S. PEIRCE SOCIETY, *Simmons*
Chair: Thomas L. Short
I. Felicia Kruse, "Semiotics and Metaphysics"
II. Thomas Olshewsky, "Peirce and the Brain-in-the-Vat Problem"
III. Peirce Essay Contest Paper
IV. Business Meeting

8. 5:15 ASSOCIATION FOR PHILOSOPHY OF THE UNCONSCIOUS
Boston University
Topic: Law
Chair: Wilfried Ver Eecke
Phillippe Van Hautte, "Law in Lacan and Foucault"
Commentator: Rick Boothby

9. 5:15 ASSOCIATION FOR PHILOSOPHY JOURNAL EDITORS
New Hampshire
Cocktail Reception for all Journal Editors to set future agenda of the association

VI. EVENING (7:30–10:30)

1. 7:30 LEIBNIZ SOCIETY OF NORTH AMERICA, *Brandeis*
Chair: Robert Sleigh
I. Christia Mercer, "Putting Mind to Matter: Leibniz's Early Conception of Substance"
II. Business Meeting

2. 7:30 SOCIETY FOR MEDIEVAL AND RENAISSANCE PHILOSOPHY, *Suffolk*
Chair: Edward P. Mahoney
I. Tamar Rudavsky, "Neo-Platonic Motifs in Medieval Jewish Discussions of Time"
II. James Hankins, "Anti-Platonism in Fifteenth-century Italy"

3. 7:30 NORTH AMERICAN KANT SOCIETY, *Radcliffe*
Chair: Karl Ameriks
Allen Wood, "Kant's Moral Anthropology"
Commentator: Jerome B. Schneewind

4. 7:30 JASPERS SOCIETY, *Wellesley*
Chair: Alan Olson
Topic: Jaspers and Heidegger
I. Tom Rockmore, "Moral Ambiguity and Jaspers's *Die Schuldfrage*"
Commentator: George Pepper

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

II. Joseph Margolis, "On the Moral and Political Responsibility of Intellectuals"
Commentator: Raymond Langley

5. 7:30 SOCIETY FOR ASIAN AND COMPARATIVE PHILOSOPHY, *Bentley*
Chair: Monte S. Hull
Topic: Evidence and Confirmation of Philosophic Claims East and West
Speakers: Purusottama Bilimoria and Robert E. Allinson

6. 7:30 PERSONALIST DISCUSSION GROUP, *Tufts*
Chair: Thomas O. Buford
Topic: Is God a Person?
Speakers: Robert Neville and Rufus Burrow, Jr.

7. 7:30 CONFERENCE OF PHILOSOPHICAL SOCIETIES, *Regis*
Topic: The Rediscovery of the American Tradition in Philosophy
Chair: John A. Loughney
Speakers: Marjorie C. Miller, Vincent Colapietro, Peter H. Hare, and John E. Smith

8. 7:30 SOCIETY FOR THE PHILOSOPHIC STUDY OF MARXISM, *MIT*
Topic: Marxism: Analytic or Dialectical
Chair: George Hampsch
I. James Lawler, "Functionalism and a Dialectical Theory of Causality"
II. John Ryder, "Marxism and a Dialectical Ontology"

9. 7:30 SANTAYANA SOCIETY, *Harvard*
Chair: Paul G. Kuntz
I. Richard Lyon, "Oliver's Last Soliloquies"
Commentator: Angus Kerr-Lawson
II. Herman Saatkamp, Jr., Critical Edition of *The Last Puritan*

10. 7:30 YOUNG PHILOSOPHERS SOCIETY FOR INTERNATIONAL EXCHANGE *Falmouth*
I. Darrell Moellendorf, "A Marxist Approach to Internationalism"
II. Laura Duhan, "Soviet Feminism"
III. Patrick Goold, "Two Czechs on Reality"
IV. Business Meeting

11. 7:30 ASSOCIATION FOR PHILOSOPHY OF EDUCATION, *Simmons*
Chair: Jane Duran
I. Robert N. Van Wyk, "Is Value Education the Achilles' Heel of Liberalism?"
Commentators: Barry Bull and R. M. Hare
II. Business Meeting

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Friday, December 28, Evening, VI

12. 7:30 RADICAL PHILOSOPHY ASSOCIATION, *Northeastern*
Topic: Philosophy and Socialism in Cuba, May 1990
Chair: Cliff DuRand
Speakers: Joy Gordon, Andrea Nye, Pat Mills, and John Torgeson

13. 7:30 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Boston University*
Topic: Justice and Gay Rights
Chair: Ed Stein
Morris B. Kaplan, "Autonomy, Equality, and Community: The Question of Lesbian and Gay Rights"
Commentators: Rosemarie Tong and Robert J. Anderson

SATURDAY, DECEMBER 29

VII. MORNING (9:00–11:00)

1. 9:00 SOCIETY FOR MEDIEVAL AND RENAISSANCE PHILOSOPHY, *Suffolk*
Chair: John A. Trentman
I. Stephen F. Brown, "Analogy in the Thirteenth and Fourteenth Centuries"
II. Stephen Dumont, "Quasi Aristotelianism and Freedom of the Will"

2. 9:00 SOCIETY FOR THE PHILOSOPHIC STUDY OF GENOCIDE AND THE HOLOCAUST, *Radcliffe*
Chair: Sander H. Lee
I. Michael A. Grodin and George J. Annas, " The Nazi Doctors and the Nuremberg Code: Human Rights in Human Experimentation"
II. Stephen Post, "Ethical Issues Generated by the Use of Data from Nazi Medical Experiments"

3. 9:00 WORLD PHENOMENOLOGICAL INSTITUTE, *Bentley*
Chair: George Williams
I. J. N. Mohanty, "Religion and the Concept of the Sacred"
II. Anna-Teresa Tymieniecka, "The Inward Sacred in the Human Being: The Enigma of Its Genesis and the Other"
III. Krystyna Gorniak-Kocikowska, "The Sacred: Inspiration or Creative Invention"

4. 9:00 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Brandeis*
Topic: Deduction and Dialectic
Chair: Robert Berman
Speakers: David Stern and James H. Wilkinson

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

VIII. AFTERNOON (2:15–4:15)

1. 2:15 AMERICAN ASSOCIATION FOR THE PHILOSOPHIC STUDY OF SOCIETY, *Brandeis*
Topic: Tibor R. Machan's *Individuals and Their Rights*
Chair: Douglas J. Den Uyl
Speakers: Paul Gaffney and R. G. Frey
Respondent: Tibor R. Machan

2. 2:15 SOCIETY FOR THE PHILOSOPHY OF CREATIVITY, *Radcliffe*
Topic: Contemporary Chemistry and Philosophy
Chair: Raymond D. Boisvert
Joseph Earley, "The Rock and the Flame: Self-Organizing, Complex Natural Systems and Their Importance for Philosophy"
Commentators: Frederick Ferre and Peter A. Y. Gunter

3. 2:15 SOCIETY FOR PHILOSOPHY AND PUBLIC AFFAIRS, *Wellesley*
Chair: Diana T. Meyers
James Nickel, "Minorities, Nationalities, and Indigenous Peoples: Can Uniform Constitutional Rights Fit Them All?"
Commentators: Bernard Boxill and Judith Lichtenberg

IX. EVENING (6:30–9:30)

1. 6:30 ASSOCIATION FOR PHILOSOPHY AND LIBERATION, *Suffolk*
Chair: Oliva Blanchette
I. Kate Lindemann, North American Perspective
II. Joseph Prabhu, Asian Perspective

2. 6:30 INTERNATIONAL SOCIETY FOR NEO-PLATONIC STUDIES, *Radcliffe*
Topic: John Scottus Eriugena
Chair: Sheri Katz
I. Dermot Moran, "Johannes Scottus Eriugena: An Early Idealist?"
Commentator: Paul Eisenberg
II. Willemien Otten, "The Role of Dialectic in the *Periphyseon's* Concept of the Return"
Commentator: Donald F. Duclow

3. 6:30 JASPERS SOCIETY OF NORTH AMERICA, *Wellesley*
Topic: Newman and Jaspers
Chair: Harold Oliver
I. Arthur Kennedy, "The University as Constituting Agent of Culture"
Commentator: Charles Courtney
II. Gregory Walters, "Academic Freedom and the University Idea"
Commentator: Malek Khazaee

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Saturday, December 29, Evening, IX

4. 6:30 SOCIETY FOR THE PHILOSOPHY OF SEX AND LOVE, *Bentley*
Chair: Carol Caraway
I. David B. Annis, "Intimacy, Friendship, and Autonomy"
Commentator: Alan Soble
II. Natalie Dandekar, "Eros, Romantic Illusions, and Political Opportunism"
Commentator: Carol S. Gould

5. 6:30 SOCIETY FOR WOMEN IN PHILOSOPHY, *Simmons*
Topic: In Honor of Distinguished Women Philosophers
Chair: Margaret Walsh
1990 Honoree: Sandra Harding
Respondents: Naomi Scheman and Linda Alcoff
Business Meeting

6. 6:30 SOCIETY FOR THE STUDY OF WOMEN PHILOSOPHERS
Boston University

7. 6:30 SOCIETY FOR THE PHILOSOPHIC STUDY OF THE CONTEMPORARY VISUAL ARTS, *Boston College*
Chair: Sander Lee
I. Thomas Heyd, "Understanding Peformance Art: Art Beyond Art"
II. Mark Roche, "Philosophic and Aesthetic Structures in Woody Allen's *Crimes & Misdemeanors*"
III. Paul Santilli "Spike Lee's Lingua Fracas: *Do the Right Thing*"
IV. Business Meeting

8. 6:30 CONFERENCE OF PHILOSOPHICAL SOCIETIES, *Regis*
Topic: Philosophy's Contributions to Counseling and Psychotherapy
Chair: James P. Cadello
Speakers: Elliot Cohen, Paul Sharkey, Brad Art, and Katherine McCall
Business Meeting

9. 6:30 SOCIETY FOR THE STUDY OF ETHICS AND ANIMALS, *MIT*
Chair: Lilly-Marlene Russow
I. Roberta Kalechofsky, "Jewish Law, Tradition, and Animal Rights: Is There a Usable Paradigm for the Movement?
Commentator: Eric Katz
II. Evelyn B. Pluhar, "The Joy of Killing"
Commentator: Daniel A. Dombrowski

10. 6:30 WORLD PHENOMENOLOGICAL INSTITUTE, *Northeastern*
Chair: Benjamin I. Schwartz
I. Ynhui Park, "The Transnatural Destiny of the Soul and the Way of the Tao"

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

II. Paul Brockelmans, "God and the Pursuit of Being"
III. Steven W. Laycock, "Telic Divinity and Its Atelic Ground"

11. 6:30 ASSOCIATION FOR INFORMAL LOGIC AND CRITICAL THINKING
Nantucket
I. *Topic:* Book Discussion: David Sanford's *If P Then Q: Conditionals and the Foundations of Reasoning*
Speakers: Robert Stalnaker and Vann McGee
Respondent: David Sanford
II. *Chair:* Bertram Bandman
Craig Walton, "Critical Thinking and the Art of Judgment"
Commentator: George Bowles

12. 6:30 RADICAL PHILOSOPHY ASSOCIATION AND APA COMMITTEE ON BLACKS, *Harvard*
Topic: The American Evasion of Philosophy
Chair: Leonard Harris
Speakers: Cornel West, Robert Gooding Williams, and Elizabeth V. Spelman

13. 6:30 SOCIÉTÉ AMÉRICAINE DE PHILOSOPHIE DE LANGUE FRANÇAISE
Brandeis

SUNDAY, DECEMBER 30

X. MORNING (9:00–11:00)

1. 9:00 SOCIETY FOR WOMEN IN PHILOSOPHY, *Wellesley*
Chair: Carol Nicholson
I. Margaret Walker
II. Kristin Waters, "Epistemology Reclaimed: Feminist Ethics and Systems of Knowledge"

2. 9:00 AMERICAN SOCIETY FOR VALUE INQUIRY, *Brandeis*
Topic: Abortion
Chair: Betsy Postow
I. Daniel A. Dombrowski, "Asymmetrical Relations, Identity, and Abortion"
II. Willem de Vries, "I'm Glad My Mother Didn't Have an Abortion"
III. Terry McGarrity, "Is Abortion Morally Permissible?"

3. 9:00 SARTRE CIRCLE, *Suffolk*
I. Ronald E. Santoni, "Sartre's Prephilosophical Rejection of God"
II. Steven W. Laycock, "Nothingness and Emptiness: Exorcising the Shadow of God in Sartre"
Commentator: Thomas W. Busch

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

GROUP PARTICIPANTS
(See Appendix)

ABBARNO, John, D'Youville College, New York II-1
ADDELSON, Kathryn Pyne, Smith College, Massachusetts II-11, III-3
AIKEN, William, Chatham College, Pennsylvania V-2
ALCOFF, Linda, Syracuse University IX-5
ALLAN, George, Dickinson College, Pennsylvania II-5
ALLINSON, Robert E., Deakin University, Australia VI-5
AMERIKS, Karl, University of Notre Dame VI-3
ANDERSON, Robert J., Washington College, Maryland VI-13
ANNAS, George J., Boston University VII-2
ANNIS, David B., Ball State University IX-4
ARNSTINE, Barbara, California State College, California IV-4
ARNSTINE, Donald, University of California/Davis IV-4
ART, Brad, Westfield State College, Massachusetts IX-8
ASH, Badwell, Suffolk University II-7
ASHBAUGH, Anne, Colgate University IV-3
AXINN, Sidney, Temple University II-9
BANDMAN, Bertram, Long Island University IX-11
BARRAL, Mary Rose, Gannon University II-4, III-7
BENSON, Peter, McGill University, Canada I-6
BERNAUER, James, Boston College, Massachusetts II-3
BERMAN, Robert, Xavier University, Louisiana VII-4
BILIMORIA, Purusottama, Chinese University of Hong Kong VI-5
BLANCHETTE, Oliva, Boston College, Massachusetts IX-1
BOISVERT, Raymond D., Siena College, New York VIII-2
BOK, Sissela, Brandeis University IV-5
BOOTHBY, Rick, Loyola College, Maryland V-8
BOWLES, George, George Washington University IX-11
BOXILL, Bernard, University of North Carolina/Chapel Hill VIII-3
BRAIDER, Christopher, Harvard University III-7
BRENNAN, Andrew, University of Stirling, Scotland III-2, V-2
BROCKELMANS, Paul, University of New Hampshire IX-10
BROWN, Stephen F., Boston College, Massachusetts VII-1
BUFORD, Thomas O., Furman University VI-6
BULL, Barry, University of Indiana VI-11
BURROW, JR., Rufus, Christian Theological Seminary, Indiana VI-6
BUSCH, Thomas W., Villanova University X-3
CADELLO, James P., Regis College, Colorado IX-8
CARAWAY, Carol, Indiana University, Pennsylvania IX-4
CARPENTER, Stanley R., Georgia Institute of Technology II-8
CASEBIER, Allan, University of Southern California II-12, V-5
CASULLO, Albert, University of Nebraska IV-1
CAVELL, Stanley, Harvard University II-10
COHEN, Elliot, Indian River Community College, Florida IX-8
COHEN, Joshua, Massachusetts Institute of Technology I-6
COLAPIETRO, Vincent, Winona State University VI-7
CONANT, James, University of Pittsburgh & University of Michigan II-10
CORRINGTON, Robert S., Drew University III-5
COURTNEY, Charles, Drew University IX-3

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

CREEGAN, Charles, North Carolina Wesleyan College II-7
CUMMINS, Phillip D., University of Iowa V-1
CUNNINGHAM, Nance, Yale School of Medicine V-4
DAHBOUR, Omar, City University of New York/Hunter College I-5
DANDEKAR, Natalie, University of Rhode Island IX-4
D'ANDRADE, Kendall, Shawnee State University II-2
DE VRIES, Willem, University of New Hampshire X-2
DEJNOZKA, Jan, Union College, New York IV-1
DEN UYL, Douglas J., Bellarmine College, Kentucky VIII-1
DETMER, David, Purdue University/Calumet III-4
DOMBROWSKI, Daniel A., Seattle University IX-9, X-2
DONNELLY, John, University of San Diego II-7
DUCLOW, Donald F., Gwynedd-Mercy College, Pennsylvania IX-2
DUHAN, Laura, University of North Carolina/Charlotte VI-10
DUMONT, Stephen, Pontifical Institute of Medieval Studies VII-1
DURAN, Jane, University of California/Santa Barbara VI-11
DURAND, Cliff, Morgan State University VI-12
DURBIN, Paul T., University of Delaware II-8
EARLEY, Joseph, Georgetown University VIII-2
EBENRECK, Sara, Prince Frederick, Maryland III-2
EISENBERG, Paul, Indiana University IX-2
ELGIN, Catherine Z., Dartmouth College, New Hampshire IV-4
FERRE, Frederick, University of Georgia VIII-2
FLEMING, Richard, Bucknell University II-10
FREDE, Dorothea, Swarthmore College, Pennsylvania V-6
FREY, R. G., Bowling Green State University VIII-1
FRIEDMAN, Marilyn, Purdue University III-3
GAFFNEY, Paul, Saint John's University VIII-1
GASPER, Phil, Middlebury College Vermont I-5
GAVIN, William J., University of Southern Maine III-5
GOLD, Steven Jay, Southern Connecticut State University II-11
GOODMAN, Nelson, Harvard University IV-4
GOOLD, Patrick, Saint Olaf's College, Minnesota VI-10
GORDON, Joy, Yale University VI-12
GORNIAK-KOCIKOWSKA, Krystyna, Temple University VII-3
GOTTHELF, Allan, Trenton State College, New Jersey III-1
GOTTLIEB, Roger, Worcester Polytechnic Institute II-11
GOULD, Carol S., Union College, New York IX-4
GOULD, Tim, Metropolitan State College, Colorado II-10
GRACIA, Jorge, State University of New York/Buffalo IV-3
GRANGE, Joseph, University of Southern Maine II-5
GRODIN, Michael A., Boston University VII-2
GUNTER, Peter A. Y., University of North Texas VIII-2
GUY, Fred, University of Baltimore IV-1
HAAS, Peter, Vanderbilt University II-3
HALL, Ronald, Francis Marion College, South Carolina II-7
HAMPSCH, George, Holy Cross College, Massachusetts VI-8
HANKINS, James, Harvard University VI-2
HANSON, Karen, Indiana University II-10, IV-4
HARDING, Sandra, University of Delaware IX-5
HARE, Peter H., State University of New York/Buffalo VI-7

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

HARE, R. M., University of Florida VI-11
HARRIS, Leonard, Morgan State University IX-12
HEYD, Thomas, University of Western Ontario, Canada IX-7
HOLVECK, Eleanore, Duquesne University III-4
HULL, Monte S., Lake Forest College, Illinois VI-5
KALECHOFSKY, Roberta, Micah Publications IX-9
KAPLAN, Morris B., State University of New York/Purchase VI-13
KATZ, Eric, New Jersey Institute of Technology V-2, IX-9
KATZ, Sheri, Spring Hill College, Alabama IX-2
KELLNER, Douglas, University of Texas/Austin II-12
KENNEDY, Arthur IX-3
KERR-LAWSON, Angus, University of Western Ontario, Canada VI-9
KHAZAEE, Malek, California IX-3
KOHL, Marvin, State University of New York/Fredonia IV-1
KOLENDA, Konstantin, Rice University III-5
KRUSE, Felicia, Xavier University V-7
KUNTZ, Paul, Emory University VI-9
KUYKENDALL, Eleanor, State University of New York/New Paltz III-4
LANGLEY, Raymond J., Manhattanville College, New York II-2, VI-4
LAWLER, James, State University of New York/Buffalo VI-8
LAYCOCK, Steven W., University of Toronto, Canada IX-10, X-3
LEE, Kwang-Sae, Kent State University II-9
LEE, Sander, Keene State College, New Hampshire II-12, VII-2, IX-7
LENNOX, James G., University of Pittsburgh III-1
LI, Xiaorong, Stanford University II-9
LICHTENBERG, Judith, University of Maryland/College Park VIII-3
LINDEMANN, Kate, Mount Saint Mary College, New York IX-1
LOUGHNEY, John A., Westfield State College, Massachusetts I-1, II-12, VI-7
LOVEKIN, David, Hastings College, Nebraska II-8
LYON, Richard, Hampshire College, Massachusetts VI-9
MACHAN, Tibor R., Auburn University VIII-1
MAGNELL, Thomas, Drew University II-1
MAHONEY, Edward P., Duke University VI-2
MAKER, William, Clemson University III-8
MARGOLIS, Joseph, Temple University VI-4
MCCALL, Katherine, Montclair State College, New Jersey IX-8
MCCORMICK, Peter, University of Ottawa, Canada V-5
MCCROSSIN, Trip, Yale University IV-1
MCGARRITY, Terry, State University of New York/Buffalo X-2
MCGEE, Vann, Rutgers University/New Brunswick IX-11
MCINERNEY, Peter, Oberlin College, Ohio III-4
MEDINA, Vicente, Seton Hall University IV-3
MERCER, Christia, University of California/Irvine VI-1
MEYERS, Diana T., University of Connecticut/Storrs VIII-3
MILLER, Marjorie C., State University of New York/Purchase VI-7
MILLS, Pat, University of Massachusetts/Amherst VI-12
MINEAU, Andre, University of Sudbury, Canada II-3
MOELLENDORF, Darrell, Riverside City College, California VI-10
MOHANTY, J. N., Temple University V-5, VII-3
MOORE, Jennifer, University of Delaware IV-2
MORAN, Dermot, University College Dublin, Ireland IX-2

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

MORRIS, Phyllis S., Oberlin College, Ohio	III-4
MULHERN, Mary, Mulhern Associates	V-4
NEVILLE, Robert, Boston University	II-5, VI-6
NEWELL, J. David, United States Naval Academy	V-4
NICHOLSON, Carol, Rider College, New Jersey	X-1
NICKEL, James, University of Colorado/Boulder	VIII-3
NORTON, Bryan, Georgia Institute of Technology	III-2
NOTA, John, Creighton University	II-4
NYE, Andrea, University of Wisconsin	VI-12
OCHS, Peter, Drew University	I-2
OLIVER, Amy, American University	IV-3
OLIVER, Harold, Boston University	IX-3
OLSHEWSKY, Thomas, University of Kentucky	V-7
OLSON, Alan, Boston University	VI-4
OTTEN, Willemien, Loyola University of Chicago	IX-2
PARK, Ynhui, Simmons College, Massachusetts	IX-10
PATTISON, George, Suffolk University	II-7
PEPPER, George, Iona College, New York	VI-4
PEREBOOM, Derk, University of Vermont	I-4
PLUHAR, Evelyn B., Pennsylvania State University/Fayette	IX-9
POST, Stephen, Case Western Reserve University	VII-2
POSTOW, Betsy, University of Tennessee/Knoxville	X-2
PRABHU, Joseph, California State University/Los Angeles	IX-1
PREUS, Anthony, State University of New York/Binghamton	V-6
RAPP, Carl, University of Georgia	III-8
RASMUSSEN, Douglas, Saint John's University	III-1
ROBINSON, Thomas M., University of Toronto, Canada	V-6
ROCHE, Mark, Ohio State University	IX-7
ROCKMORE, Tom, Duquesne University	VI-4
ROLSTON, Holmes, Colorado State University	III-2
ROSE, Mary Carmen, Goucher College, Maryland	I-2
ROSENBERG, Alan, City University of New York/Queens College	II-3
RUDAVSKY, Tamar, Ohio State University	VI-2
RUSSOW, Lilly-Marlene, Purdue University	IX-9
RYBA, Thomas, Purdue University	III-7
RYDER, John, State University of New York/Cortland	VI-8
SAATKAMP, JR., Herman, Texas A & M University	VI-9
SAGOFF, Mark, University of Maryland	III-2
SANFORD, David, Duke University	IX-11
SANTILLI, Paul, Siena College, New York	IX-7
SANTONI, Ronald E., Denison University	X-3
SAWICKI, Jana, University of Maine/Orono	III-3
SCHEFFLER, Israel, Harvard University	IV-4
SCHEMAN, Naomi, University of Minnesota	IX-5
SCHMITT, Richard, Brown University	I-5
SCHNEEWIND, Jerome B., Johns Hopkins University	VI-3
SCHUTTE, Ofelia, University of Florida	IV-3
SCHWARTZ, Benjamin I., Harvard University	IX-10
SHARKEY, Paul, University of Southern Mississippi	IX-8
SHARP, Gene	IV-5
SHORT, Thomas L., Kenyon College, Ohio	V-7

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

SILLS, Chip, United States Naval Academy III-8
SLEIGH, Robert, University of Massachusetts/Amherst VI-1
SMITH, David, University of California/Irvine V-5
SMITH, John E., Yale University VI-7
SOBLE, Alan, University of New Orleans IX-4
SOMMERS, Cristina H., Clark University III-3
SPELMAN, Elizabeth V., Smith College, Massachusetts IX-12
STALNAKER, Robert, Massachusetts Institute of Technology IX-11
STEIN, Ed, Massachusetts Institute of Technology VI-13
STEINBERG, Eric, City University of New York/Brooklyn College V-3
STERN, David, University of Toledo VII-4
TONG, Rosemarie, Davidson College, North Carolina VI-13
TORGESON, John, Drake University VI-12
TRACY, Thomas, Bates College, Maine I-4
TRAIGER, Saul, Occidental College, California V-3
TRENTMAN, John A., Huron College, South Dakota VII-1
TYMIENIECKA, Anna-Teresa, World Phenomenology Institute VII-3
VAN CLEVE, James, Brown University V-1
VAN HAUTTE, Phillipe, University of Leuven, Belgium V-8
VAN WYK, Robert N., University of Pittsburgh/Johnstown VI-11
VER EECKE, Wilfried, Georgetown University V-8
WALKER, Margaret, Fordham University X-1
WALSH, Margaret, Bradford College, Massachusetts IX-5
WALTERS, Gregory, Saint Mary's University IX-3
WALTON, Craig, University of Las Vegas/Nevada IX-11
WARTOFSKY, Marx, City University of New York/Baruch College II-8
WATERS, Kristin, Clark University X-1
WATTLES, Jeffrey, Kent State University II-9
WERNER, Richard, Hamilton College, New York IV-5
WEST, Cornel, Princeton University III-5, IX-12
WILKINSON, James H., University of Kentucky VII-4
WILLARD, Dallas, University of Southern California V-5
WILLIAMS, George, Harvard Divinity School VII-3
WILLIAMS, Robert, Hiram College, Ohio III-7
WILLIAMS, Robert Gooding, Amherst College, Massachusetts IX-12
WINFIELD, Richard D., University of Georgia I-6
WINKLER, Kenneth P., Wellesley College, Massachusetts V-1
WOOD, Allen, Cornell University VI-3
ZELECHOW, Bernard, York University, Canada II-7

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVII, NO. 11, NOVEMBER 1990

THE PRAGMATIST SIEVE OF CONCEPTS: DESCRIPTION VERSUS INTERPRETATION*

Two incompatible vocabularies, based on two incompatible sets of assumptions, can be found throughout pragmatist writings. The appeal to unmediated facts and induction is a remnant of Aristotelian metaphysics and Baconian science, while appeals to interpretation and translation draw on nineteenth-century hermeneutics and Darwinian science. Consequently, pragmatism has sometimes been categorized as a version of realism and sometimes as idealism. But the originality of pragmatism cannot be fully grasped without recognizing that it is neither, because its analysis of concrete human experience rejects the metaphysical assumptions on which the distinction is based.

William James's metaphorical explanations of how we find ourselves within the world are tantalizingly ambiguous. In *The Principles of Psychology,*[1] for instance, he compares a conceptual scheme to a sieve used to gather up the contents of the world: "Most facts and relations fall through its meshes, being either too subtle or insignificant to be fixed in any conception. But whenever a physical reality is caught and identified as the same with something already conceived, it remains on the sieve" (PP, I, 455). But if facts and relations are falling through our conceptual nets, how can we even talk about them? The pragmatists deny that the limits of our understanding coincide with the limits of the world. Since they also his-

*To be presented in an APA symposium on the Centennial of William James's *Principles of Psychology,* December 28. Gerald E. Myers will be co-symposiast, and H. S. Thayer will comment; see this JOURNAL, this issue, 593–9 and 600–1, respectively, for their contributions.

[1] In *The Works of William James, 3 vols.* (Cambridge: Harvard, 1981). Abbreviated as PP with volume number and page.

0022-362X/90/8711/585–592

© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

toricize and contextualize human understanding to varieties of understandings, this pluralistic proliferation of interpretations means that different sieves will capture different facts and relations. We can undermine the hegemony of any particular way of capturing reality by comparing it to realities caught through other interpretations. But it sometimes seems that the pragmatists also compare what is caught with what remains in the sea.

In *Pragmatism without Foundations,* for instance, Joseph Margolis[2] argues that "in rejecting foundationalism and conceding the historicity of human inquiry" we can no longer maintain a sharp demarcation between realism and idealism (*ibid.*, p. 203). Nonetheless, he wants to retain scientific realism because it alone supports the claim that there is a "common world that we are trying to understand and interpret through our various theories" (*ibid.*, p. 300). Moreover, he believes that we have genuine "knowledge *of* the mind-independent world (though not as it is mind-independently . . .), but only as textualized" (*ibid.*, p. 283).

Like the classical American pragmatists, Margolis wants to reconcile the realist belief in an external world and the relativist rejection of transparent cognitive access to it. But the term 'scientific realism' seems too historically overdetermined to express the pragmatist sense that the commonality of the world is not a given factor but is always at risk. It must be continually reaffirmed, developed, communicated, and accepted by particular historically evolved social groups, scientific and humanistic disciplines, political organizations, etc. Scientific realism can assume that a common world exists behind and is revealed in our various interpretations, but it cannot demonstrate anything beyond the fallibilist pragmatist position that the real world is the world experienced and tested within historically specific communities of interest.

But the earlier pragmatists are responsible for some of the present confusion because they sometimes make the same, common-sense distinction Margolis does between the world as it is and as interpreted. On this occasion of the hundredth anniversary of *The Principles of Psychology,* I limit my remarks to James. I first briefly sketch his analysis of the concrete human condition. This calls into question the meaning-content distinction which sets a realist understanding of description against an idealist understanding of inter-

[2] New York: Blackwell, 1986.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

pretation. I then make explicit the implicit move away from a traditionally empiricist notion of description to a pragmatically rational one, and show how a pure constructivist notion of interpretation is denied in favor of an interactively constructivist one.

James begins with "the world's concrete fulness," the reality of which overflows our purposes in organizing it "at every pore" (PP, II, 962, n11; 961). But such a world is not set over against us as an independent entity. We find ourselves always already within and continuous with the world. "Without selective interest, experience is an utter chaos," because we can only focus on and interact with a limited number of aspects out of the encompassing horizon (PP, I, 381).[3] But we do not experience such chaos because the human organism is selective on every level, from nonconscious physiological screening processes to highly abstract, conscious ones.

The holistic constitution of experience is also expressed in the often repeated phrase that 'all our knowledge at first is vague' (PP, II, 969). But this vagueness varies from an overwhelming much-at-onceness to the focused elimination of extraneous material typical of laboratory science. The "great blooming, buzzing confusion" felt by "the baby, assailed by eyes, ears, nose, skin, and entrails at once" soon develops into the ordered world of everyday experience (PP, I, 462). Only by reflection can we begin to retrieve the background conditions that are operative. This is not a retrieval of chaos but of the structures that order experience. Since the process of dissociation of particular objects from the fullness of experience soon becomes habitual, we easily lose sight of the fact that we continue to participate in the interactive constitution of the world.

We always already find ourselves in a world with which we have to do, so James takes activity as the model for how facts come to be. Thinking presupposes this prior activity, which is why the reflex arc, restructured to recognize the spontaneity of the organism, precedes his explanation of reasoning in *Principles*. Reasoning is a conscious process of extracting partial characters out of a phenomenal totality according to our interests (PP, II, 959ff). Since not all the possible characterizations of phenomena can be abstracted at once, or even at all, some criteria must be used to distinguish those relevant aspects

[3] For the problematic status of a primitive chaos in James, see my "Hodgson's Influence on James's Organization of Experience," in Corrington, Hausman, and Seebohm, eds., *Pragmatism Considers Phenomenology* (Washington: Center for Advanced Research in Phenomenology, 1987), pp. 187–202.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of phenomena which earlier philosophical traditions called essences. James argues that, in the full concrete act of thought, we order sensations both aesthetically, i.e., according to rules of simplicity, unity, and identity, and practically, i.e., as means to ends in order to act. Since we need a basis for choosing one course of action over another, moral principles are included among these practical interests.

James explains such traditional characterizations of the human as language, humor, and even reflective self-knowledge as refinements of the basic act of dissociation from a phenomenal totality (PP, II, 979–983). The ability to dissociate, in turn, depends on the imaginative power to associate analogically. Without the ability to dissociate, we would remain "swamped in the total phenomenon which [we] help constitute." But without the ability imaginatively to bring together phenomena widely separated in space and time or immersed in nonsimilar characteristics, our intellect could only be reproductive, not productive. Human intellection is not distinguished by our ability to reproduce phenomena but imaginatively to produce them.

Our active involvement with a world that appears to be simply given can only be reflectively realized. James says in his last work[4] that we cannot tell about what we now hold in our hands and read how much comes in through our eyes and fingers and how much from our apperceiving intellect. He thus reaffirms what he had earlier said in *Principles:* "The sign and the signified melt into what seems to us the object of a single pulse of thought. . . . In these first and simplest inferences the conclusion may follow so continuously upon the 'sign' that the latter is not discriminated or attended to as a separate object by the mind" (PP, II, 953–4).

James credits empiricism with reflectively justifying "our instinctive feeling about immediate experience," but it is his own hermeneutics of the "*full* fact" which has done so (SPP, 59). Reality should be understood from the perspective of the "*full* fact," i.e., as "a conscious field *plus* its object as felt or thought of *plus* an attitude towards the object *plus* the sense of a self to whom the attitude belongs."[5] The object taken alone is a mere abstract element of this fuller experience. His radical empiricism criticizes both the reifica-

[4] *Some Problems of Philosophy,* in *The Works of William James* (Cambridge: Harvard, 1979), p. 58. Abbreviated as SPP.

[5] *The Varieties of Religious Experience,* in *The Works of William James* (Cambridge: Harvard, 1985), p. 393.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tion of concepts that is the legacy of Platonic idealism and the isolation of bare sense data as the building blocks of reality, which is the legacy of Humean empiricism.

"The manyness-in-oneness which perception offers is impossible to construe intellectually" (SPP, 51). James understands conceptualization, from single meanings to the thematization of scientific theories, as constructs in an intellectual arrangement that facilitates the elegant or expeditious prediction of actual events (SPP, 103, n10). Like tools and habits, they are ways of organizing the flux of experience to drive it better for our ends. This active transformation of concrete experience is rhetorically expressed as "interpretation" and "translation." We transform the sense order into rational equivalents by referring sensible phenomena "to their interpretants in the shape of concepts" (SPP, 42). These interpretants are prearranged into aesthetically organized schemas. Contrast this interpretive approach, based on a concrete analysis of the way we weave the chaos of impressions into an ordered world, with an appeal to facts that James himself inconsistently understood as an independent check on such interpretations.

James takes his methodological clue from the sciences of nature, which replaced the purely rationalistic criteria of simplicity and systematic coherence favored by philosophers with the evidence that could be gathered in direct observation. This insistence on submitting purely aesthetic criteria to the court of facts gathered from observation predates Edmund Husserl's motto: "To the things themselves!" In his 1900 preface to the Italian edition of *Principles,* James emphasized his intention to eliminate metaphysical questions as inappropriate to a natural science and to confine his work to describing experience such that it "could be immediately verified by everyone's own consciousness" (PP, III, 1483). Certainty could thus be achieved, since such description was thought to be free of hidden philosophical interpretations.

A neutral description of the facts of experience will provide a firm basis for philosophical interpretation. But this positivist project is doomed from the start because of what James includes in the facts observed. He wants to replace the purely intellectualist, scholastic philosophical psychology with one that seeks to do justice to our impulses and passions and "to the cooperation of our will with our intellectual life" (PP, III, 1483). His project of disentangling psychology from metaphysics involved adopting a Darwinian biological point of view. But this very point of view undermines the empiricist

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

assumption of the passivity of the senses and vindicates the role of personal spontaneity in organizing the materials given to thought.

The problem can now be clearly stated. James adopts a natural-history methodology to discover the facts of experience that have been distorted by the illicit intrusion of metaphysical assumptions. But what this method discloses is that experience is always already unified through the interests and passions of our active spontaneity. Description of facts cannot precede their interpretation because every description already involves active interpretation. But without an appeal to the facts of the case, how are we to determine which interpretations truly reflect reality and which distort it? The pragmatist answer is that interpretations do not reflect reality but are a means of utilizing experience congruent with our interests. Interests have been formalized in the various sciences and also operate unreflectively in everyday experience. When interpretations literally work to bring about the intended reorganization of the haphazard organization of experience that we have inherited, then we can recognize this success by saying euphemistically that they worked because they grasped reality.

Even in his natural-history explanations James begins with "the full concrete act of thought" and then shows how the everyday world we take for granted arises by a process of dissociation (PP, II, 982ff). To account for our abstract notion of space, for instance, he develops a genealogical account to explain how "the vague form or *quale* of spatiality" gets subdivided by consciousness (PP, II, 786). The explanation is genealogical because it traces the development of an ordered world of space out of a multiplicity of particular spaces by identifying the interests at work. In the world of everyday experience, we introduce as much order into the vastness as is sufficient for the task at hand. The reflective reconstructions of philosophers and scientists demonstrate interests that differ from everyday ones.

Primitive perceptions of space must be woven into the ordered space apprehensible to the mind by "a rather complicated set of intellectual acts" (PP, II, 787). The primordial sensational largeness must be subdivided by consciousness and added together in a synthesis to yield objectively real space. But James also privileges the accuracy of sensational immediacy. He explicitly denies that he is a constructivist and attacks Kant's synthesis of phenomena as misguided. But what *do* the senses give and what is wrong with Kantian synthesis?

It is apparently a question of origins. James says that space, for Kantians, "is a super-sensational mental product" (PP, II, 903). He

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

instead follows Berkeley's theory of vision, which begins with primitively chaotic spatial sensations, and Locke's precept that the mind can form no new simple idea that was not already in sensation (PP, II, 900–1). The antecedents of conscious spatial quality are sensational qualities, cerebrally based (PP, II, 908). Does space consciousness enter with these first optical sensations or not? James says that it does, but vaguely rather than clearly. He recalls his earlier definition of "the psychologist's fallacy," when he notes Shadworth Hodgson's warning against confusing the analysis of an idea with the means of its production (PP, II, 910–1). Mere acquaintance with space should not be confused with explicit knowledge of spatiality.

James dismisses the "Kantian machine-shop" as being mythological because it requires two consciously distinguishable moments: multiple, primitive sensations superseded by a mental production of space (PP, II, 905). Passive reception is not followed by active constitution for James because what is seen is already ordered. But he is wrong in thinking that the Kantian explanation assumes a temporally serial ordering or that it can be disproved by the fact that he has "no introspective experience of mentally producing or creating space" (PP, II, 905). He is following in Kant's footsteps, not disputing him, when he says that the higher parts of the mind "inweave the space-sensations with intellectual relations." He does break away in denying that a unitary apprehension of a single space is given to consciousness, and in arguing that our particular intuitions of multiple spaces can only with effort be made eventually to yield an abstract notion of space as such. Since our first chaotic impressions are sensational "through and through," we do not need any synthesis or "mysterious 'mental chemistry' " to create elements not present in the original data of feeling (PP, II, 838). But we do need to "make a continuous order of what was a rather incoherent multiplicity."

James wants both to deny any idealism that holds that we construct our experienced worlds out of nothing or by simply imposing categories on compliant sensations, and to deny any empiricism that holds that we build up an ordered world out of the discrete data of sensations by some mechanism of association that is the same for everyone. He starts neither with categorial analysis nor with sense data, but with our being-in-the-world. His basic hypothesis is that we weave chaos into order to satisfy aesthetic and practical interests.

Although we all uniquely order our experiences, we do not do so haphazardly. We are constrained, though we are not rigidly determined, by the particular configuration the human biological organ-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ism has developed over time. What we take to be a normal point of view that reveals to us the world as it objectively is turns out to be the one that offers the most "aesthetic and practical advantages" (PP, II, 871). From this privileged perspective, we distinguish between what is and what only seems to be. But we soon learn through experience and custom how seeming and being pass one into the other by continuous gradations. We learn to translate an object perceived from an abnormal perspective into the real object, i.e., we soon distinguish the sign from the reality. But this can only be reality-for-us because we are implicated, both physiologically and culturally, in the determination of normalcy and of the criteria for objectivity.

The interpretative approach reflects the pragmatist position that we find ourselves always already within a world that we help constitute. The descriptive approach should therefore be recast in its terms rather than reduce interpretation to some pure description of reality. This is not a free-floating constructivism, however, but one that seeks to take account of facts as experienced by tracing them to their informing interests. The false metaphysical assumptions behind the simple appeal to given facts should therefore be systematically identified and rejected. James rephrases his reconstructed reflex-arc thesis, which holds that we are interactively involved in the world, as a hermeneutic approach, which explains how this activity influences our understanding of the world: "The more we see, the more we think; while the more we think, the more we see in our immediate experiences, and the greater grows the detail, and the more significant the articulateness of our perception" (SPP, 59).

CHARLENE HADDOCK SEIGFRIED

Purdue University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

JAMES AND FREUD*

During this centennial celebration of William James's *The Principles of Psychology,*[1] others are doubtlessly voicing the sentiments of a contemporary psychologist: "*Principles* . . . is the single greatest work in American psychology. Among books written by psychologists, its only rival is Sigmund Freud's *The Interpretation of Dreams.*"[2] Such compliments to James often provoke comparisons with Freud, since they are commonly linked in being called the two "titans" of modern psychology, and because both explored human nature through the study of abnormal behavior. Not to be forgotten is the amount of pathological material, accompanied by James's theorizing about consciousness and the unconscious, that occurs in *Principles.*

James and Freud met once, in 1909, when Freud delivered his *Five Lectures on Psycho-Analysis* at Clark University. Ernest Jones[3] reported that James said to him there, alluding to the growing interest in the unconscious that Freud's writings had aroused: "The future of psychology belongs to your work" (*ibid.,* p. 261). References by James and Freud to each other are few, and, while generally respectful, they indicate, especially from Freud's side, a minimal interest in each other's contributions. James disliked the dogmatism that he found in Freud's dream symbolism and antireligiosity but commended his insistence on the reality of unconscious mental processes.

There is no evidence that either directly influenced the other. James's theorizing on the unconscious was indebted to A. Binet and P. Janet in France, to E. Gurney and F. W. H. Myers in England; any mention of Freud was merely that. It does appear that Freud was (indirectly) aware of James's chapter on instinct in *Principles.* James had drawn upon predecessors for formulating what may be called two "laws" about instincts—instincts are inhibited by habits and their duration is transient. These were adapted by a German authority on sexual pathology, Albert Moll, in an 1897 book that Freud

*To be presented in an APA symposium on William James' *Principles of Psychology,* December 28. Charlene Haddock Seigfried will be co-symposiast, and H. S. Thayer will comment; see this JOURNAL, this issue, 585–592 and 600–1, respectively, for their contributions.

[1] In *The Works of William James* (Cambridge: Harvard, 1981).

[2] Joseph Adelson, "Still Vital After All These Years," *Psychology Today,* XVI, 4 (April 1982), p. 52. Adelson's article is a review of the 1981 publication of *The Principles of Psychology.*

[3] *The Life and Work of Sigmund Freud,* Lionel Trilling and Steven Marcus, eds. (New York: Anchor, 1963).

0022-362X/90/8711/593–9
© 1990 The Journal of Philosophy, Inc.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

apparently read with care. But how Freud responded to the specifically Jamesian ideas remains shrouded.[4]

Similarities between the two are of course discoverable, and probably the best analysis of how James anticipated Freudian hypotheses is David Shakow and David Rapaport's.[5] These authors argue persuasively that James's adoption of Darwinism and development of a functional dynamic psychology, similar in such respects to Freud's, contributed to American receptivity to Freudianism. "Sexual function" is the only mention of sex in *Principles'* index and that refers to a single page in chapter two on brain functions. Nevertheless, despite some prudishness, James dropped enough interesting (unindexed) remarks about sex to convince Shakow and Rapaport, among others, that, in *Principles* and other places, James was closer to Freud than is generally realized.

Dissimilarities between the two are more conspicuous. Both were given to speculation, but Freud's was generally in the name of science whereas James's was more often in the service of intuition, psychical research, and religion. Freud sought a system of psychotherapy but James preferred personal guidelines for self-help. James, I think, would typically replace Freudian "answers" with open-ended "suggestions" for you and me to develop according to our own lights. An illuminating contrast between James and Freud in these and other respects is provided in Peter Gay's[6] recent study.

My focus here, however, is narrower. I want to try to adjudicate an *implicit* dispute between James and Freud which shows through their explicitly contrasting responses to Janet's concept of the unconscious. Both Janet and Freud had studied with Jean Charcot, but their relationship, after a congenial beginning, soured amid mutual recriminations, due largely to Janet's taking credit for Freudian ideas and accusing Freud of plagiarism. It was prior to this, of course, that James's *Principles* appeared with its substantial reliance upon Janet's work; yet it is interesting that James, almost certainly unaware of their mutual hostility, continued to endorse Janet, whom Freud increasingly regarded as a personal and professional enemy.[7]

The professional divide occurred in explaining hysterical symptoms, especially those associated with multiple-personality disorders. What intrigued James were the case histories described by Janet and

[4] See, for an informative discussion of this, Frank J. Sulloway, *Freud, Biologist of the Mind* (New York: Basic, 1979), pp. 266, 290–1.

[5] "The Influence of Freud on American Psychology," *Psychological Issues*, IV, 1, Monograph 13 (New York: International Universities Press, 1964).

[6] *A Godless Jew: Freud, Atheism, and the Making of Psychoanalysis* (New Haven: Yale, 1987), pp. 21–4. See also, Gerald E. Myers, *William James: His Life and Thought* (New Haven: Yale, 1986), pp. 167–8.

[7] Freud, *An Autobiographical Study* (New York: Norton, 1952), pp. 56–7.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

others which seem to prove the coexistence (not merely alternation) of several personalities within one individual. Assuming a familiarity with such phenomena, also with the claim that they may occur in some degree in all of us, we can look at how the James/Janet explanation of such phenomena differs from Freud's.

Roughly summarized, Janet held that, when a person suffers from a constitutional weakness that prevents him from holding together a collection of mental processes in a unified consciousness, that will cause a constriction of consciousness and a resultant "splitting" or dissociation of consciousness, permitting those split off to become secondary, tertiary, etc., personalities. If the normal primary consciousness is hypnotically or hysterically anaesthetized, then, as James put it, "during the times of anaesthesia, and coexisting with it, *sensibility to the anaesthetic parts is also there, in the form of a secondary consciousness* entirely cut off from the primary or normal one, but susceptible of being *tapped* and made to testify to its existence in various odd ways."[8]

Freud had four objections to Janet's view, and by implication to James's. He questioned the mental weakness of hysterics alleged by Janet to cause consciousness to split, because too many hysterics exhibited not weakness but the opposite: mental agility and efficiency. Secondly, once we posit a secondary consciousness, where do we stop? An infinite series of such may be generated. Thirdly, and this weighed most heavily in Freud's 1915 essay "The Unconscious," clinical evidence seemed to indicate that latent unconscious mental processes lack the essential attributes of consciousness. Fourthly, and this is our focus, anyone who argues like Janet (and James), Freud said, that "one's own hidden processes belonged actually to a second *consciousness* would be faced with the concept of a consciousness of which one knew nothing, of an 'unconscious consciousness'. . . ."[9] Instead, Freud proposed that repression, not mental splitting or dissociation, causes hysterical symptoms, and the unconscious receiving the repressed material is a *mental* yet *impersonal* system unpopulated by secondary personalities or submerged centers of consciousness. Freud suggested that the James/Janet concept of a buried secondary personality coexisting with an awake normal one implies an "unconscious consciousness," a seemingly absurd notion. But James (partially following Janet) held that, if unconscious thoughts, feelings, sensations, and memories exist, they presumably belong to *someone,* to a consciousness of some sort. This

[8] *Principles,* I, 201; see also chs. VIII, IX, X, and XXVII.
[9] *An Autobiographical Study,* p. 59; also, *Five Lectures on Psycho-Analysis* (New York: Norton, n.d.), pp. 21–8; and see, Richard Wollheim, *Sigmund Freud* (New York: Viking, 1971), pp. 187ff.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

figures in the oft-discussed claim of "The Stream of Thought" chapter in *Principles* that every thought tends to appear as part of a *personal* consciousness.

He wrote 'tends to appear' because of the need to argue that, as with multiple personalities, such thoughts that do not appear as part of one's consciousness do in fact appear at a different level of consciousness which is "unconscious" so far as ordinary, waking consciousness is concerned. Interestingly, James did not regard the proposition that thoughts are personal as being definitionally true, allowing that, whether "anywhere . . . there be a mere thought, which is nobody's thought, we have no means of ascertaining, for we have no experience of its like."[10] Still, the concept of a fully anonymous thought clearly challenged his credulity as much as an "unconscious consciousness" challenged Freud's.

Freud's concern that an "unconscious consciousness" is oxymoronic reflects the problem of finding an adequate language to fit the facts. The facts, minimally stated, are these: persons suffering from hysteria or multiple-personality disorders may be visually anaesthetized, for example, in being unable to report seeing an object before them; so far as their awake conscious selves are concerned, the object is absent. Yet, if appropriate subliminal suggestions are made to them but evidently unregistered by their ordinary consciousness, the hands of such persons, James noted, will write "words which their eyes were vainly endeavoring to 'see,' i.e., to bring to the upper-consciousness. Their submerged consciousness was of course seeing them, or the hand could not have written as it did" (*ibid.*, I, 203–4).

Describing such persons as "unconsciously seeing" or being "unconsciously aware" is, unlike "unconsciously conscious," not immediately oxymoronic. That we are all constantly aware of more than we can at the time consciously register or be consciously aware of, I suppose, is probably true. An experience includes more detail than does the consciousness of that experience, meaning by 'an experience' a set of simultaneous events that are psychological, not simply physiological. Hence, merely by substituting 'awareness' for 'consciousness' we eliminate in 'unconscious awareness' the oxymoron that troubled Freud.

But other tangles remain. What is meant by 'conscious awareness' once we admit unconscious awareness? Absent any contrast, it would seem oddly redundant. It is insufficient, though it is occasionally proposed, to equate conscious awareness with being capable of reporting, telling oneself, etc., that one is aware of something. A hysteric, "automatically" repeating or echoing what his doctor tells him,

[10] *Principles*, I, 221.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

may truly say 'I see *A*', although its truth follows not from his consciously but from his unconsciously seeing *A*. If one is not consciously aware of *A*, one's report cannot compensate for its absence. We normally report what we are conscious of, but that leaves consciousness independent of being reported.

We need to see when 'conscious awareness' is and is not redundant. If we point, as it were, to a state of awareness, an awareness of *A*, then, by definition, *that* state is *conscious* of *A*; by definition, it cannot be unconscious of *A*. Considering states of awareness as things in themselves, which we can point to individually, we see that they never accommodate the conscious/unconscious distinction, and of *them* 'conscious awareness' is always redundant.

States of awareness/consciousness can only be further described as conscious or unconscious *relative to a subject, person, or reporter of some sort,* thereby giving nonredundant sense to 'conscious awareness' and any sense at all to 'unconscious awareness/consciousness'. An individual suffering multiple-personality disorder is consciously aware of seeing *A* if his state of visual awareness of *A* is normal, i.e., if he is consciously aware of *A* and if the normal ways of reporting and discussing it are operative. He is "unconsciously" seeing *A* if his experiences include visual awareness of *A* but he can only report and discuss it via abnormal techniques such as hypnotism and subliminal suggestion, his so-called "normal" self being somehow debarred from reporting and discussing it.

If we accept the evidence as summarized earlier for occurrences of unconscious awareness, we must admit with James that a state of awareness of which one's normal self is unconscious is itself an awareness of *A*. We may also state, for James (as against Freud's claim that this faces us with a "concept of a consciousness of which one knew nothing"), that we assert the occurrence of an unconscious awareness (consciousness) only because of what abnormal techniques show its salient features to be like. Our argument takes the concept of "being aware/conscious of," whether as applied to states of awareness or in that same respect as applied to persons, as being primitive or unanalyzable. I believe that is how it is and must remain. Explaining how persons can be consciously or unconsciously aware does not explain consciousness/awareness itself. The various modes of consciousness (seeing, hearing, feeling, etc.) are known by acquaintance only, and, while endlessly describable, they are not reducible to something putatively more basic.

A question remains: How to characterize awareness states of which one's normal consciousness is unaware? Are *they* aware of *A* when the person whose states (in some respect) they are is only unconsciously aware of *A*? Or must such states belong (in another respect)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

to a "secondary" person who is consciously aware of *A*? Freud rejected these options, because if a person is not conscious of *A*, there are no unconscious states of awareness of *A*; the unconscious excludes awareness. Like Jean-Paul Sartre and others, I detect a problem here for Freud. If a "censor" acts in the unconscious to repress potentially distressing material, as Freud claimed, how does it recognize what to repress unless *it* is conscious of that material? If so, we have an unconscious consciousness resembling that of James and Janet. We no longer have just two colliding psychical systems, one conscious and the other not, as Freud held in opposing Janet, since awareness/consciousness seems to appear in both.

Nor do I grasp, though the fault is perhaps mine, how Freud can avoid "consciousness at the unconscious level" in explaining multiple-personality phenomena. In calling them secondary personalities or submerged centers of consciousness, James confronted more directly than Freud did the phenomena that still puzzle psychology; again, what is especially provocative is the apparent *coexistence* in the same individual of two or more "sites" of awareness, these being unaware of each other and of what each other is aware of.

However multiple personalities are explained, the problem of conceptualizing their symptoms should be decided by the look of the empirical evidence, not by a priori arguing that states of awareness require the built-in presence of an *I*. The evidence cited by James (as well as contemporary confirmations, including split-brain phenomena) surely upset the assumption that an individual's simultaneous states of awareness must belong to a single *I*. The claim that an "unconscious awareness" must belong to an *I* (a view to which James, despite some uncertainty, was definitely inclined) is far from clear. What is meant by 'belonging to'? How are the shifting references of 'I' to be interpreted?

We need to remember that thinking or talking via the first-person pronoun is a sophisticated achievement; complex physiological and psychological conditions must be met for such self-reference to occur. Given the fragile complexity of the human organism, that such conditions sometimes fail is hardly a startling supposition. Nor is our supposing that there are states of awareness that are both unconscious and *I*-less so startling; even less so, of course, when we cite our own experiences as evidence for such awareness. I mean (quite apart from the perplexing multiple-personality phenomena) the kinds of experiences where one becomes suddenly at *t* conscious of the fact that unconscious seeing, hearing, or feeling *A* is occurring. The introspectable "feel" of many, if not all such experiences, is reported not by 'I was aware before *t* of *A* all along' but by 'It was like a sudden discovery of something occurring in my experience,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

seeing *A*'. What one may discover, I have urged, is an already occurring awareness of *A* that until *t* had eluded consciousness.

It is disappointing that James was content to say about what I call *I*-less states of consciousness what I quoted earlier, only that "whether . . . there be a mere thought, which is nobody's thought, we have no means of ascertaining, for we have no experience of its like." It disappoints that he was not more inquisitive introspectively about what it is like to become conscious of unconscious awareness. Had he been so disposed, he might have endorsed the line of thought advocated here and have opposed both Freud's exclusion of unconsciousness from the conscious system and his rejection of secondary centers of consciousness in explaining multiple-personality oddities.

Such introspective reticence is untypical of *Principles,* which abounds in almost heroic efforts to bring subsurface experience into verbal view. Highly personalized and often strange-sounding introspective formulations give *Principles,* especially the chapters on the consciousness of self and the stream of thought, its special flavor. It takes courage to make personal, introspective accounts the basis for philosophical/psychological inferences, for those accounts are not easily formulated, requiring both linguistic and conceptual innovations, and consequently, in employing tropes, run the danger of seeming idiosyncratic and peripheral to a public discourse. To say this is to appreciate anew why *Principles* and Freud's *The Interpretation of Dreams* invite comparison.

On the whole, however, James's are heavy-duty introspections as compared to Freud's. The motivation and diagnosis of aberrant ideas and behavior intrigued Freud, and, apart from his own self-diagnosis, macroscopic theory construction more than microscopic introspection nourished his hope for a science of psychoanalysis. But for James, neither motivation nor diagnosis, nor science, represented the real goal. *That,* I think, was to show how it is that psychology, unlike the natural sciences, never escapes the individual and personal. For James, unlike Freud, introspections that accentuate the "twang" of sensations were less trivial than those which serve primarily as mere data for generalizations. Whereas for others the aesthetic and literary flavor of personal introspections are psychology's nemesis, for James a psychology without such flavor was as useless as it was tasteless.

GERALD E. MYERS
City University of New York/Queens College

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

COMMENTS ON SIEGFRIED AND MYERS*

Gerald E. Myers gives us a sensitive and often subtle discussion of certain difficulties in interpreting relations between unconscious and conscious mental states. He finds a fundamental difference between William James and Sigmund Freud on this matter. The historical setting enables us to appreciate the gravity of the problem—as addressed by two very great thinkers—and some of its perplexing and elusive features.

Both James and Freud regarded thought as "part of a personal consciousness." (Janet, incidentally, held that there were exceptions to this assumption.) Both affirmed the momentous significance for psychology of unconscious states and processes. As Myers points out, however, James viewed some occurrences of dissociated personality as constituting plural centers of consciousness, or "personages," occupying unconscious strata of a subject's "normal" consciousness. Freud repudiated the notion of conscious centers in the unconscious as paradoxical.

Myers skillfully uncovers serious difficulties in each position. Freud's principle of the censor—so fundamental to psychoanalytic theory according to Freud—involves attributions of conscious properties to an unconscious mechanism in the unconscious. Thus a dilemma in its way reminiscent of the notorious pineal gland problem. James's view leads to conceptual complications concerning the "self," personal identity, and the semantics of the first-person singular pronoun.

From this conceptual *Götterdämmerung,* Myers proposes that some enlightenment might come through the use of 'awareness'. 'Unconscious awareness' avoids the paradoxical 'unconscious consciousness' rejected by Freud. Myers's effort is commendable, but it incurs diminishing returns as he proceeds to show that the fundamental problems remain substantially unresolved.

I suggest that 'awareness' is slippery with shifting senses; that 'unconscious awareness' is not much of a gain over 'unconscious consciousness'; and the cognate, 'conscious awareness', threatens to be tautological.

Charlene Haddock Seigfried argues for a version of pragmatism that cannot be identified as, or subsumed under, either realism or

* Abstract of a paper to be presented in an APA symposium on the centennial of William James's *Principles of Psychology,* December 28, commenting on papers by Charlene Haddock Seigfried and Gerald E. Myers; see this JOURNAL, this issue, 585–592 and 593–9, respectively, for their contributions.

0022-362X/90/8711/600–1 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

idealism. It is rather "an analysis of concrete human experience." ('Concrete' is a term of approbation recurring in her discussion but not elucidated.) Appropriate to this occasion, she develops the theme in the context of James's thought.

I am sympathetic with much of Seigfried's essay, especially her affirmation and discerning exposition of James's view of aesthetic, normative, and practical interests influencing the conceptual interpretation of the "facts" of experience. In discussing and defending the merits of this view, she comes to a question of the objectivity of interpretation. She writes: "Description of facts cannot precede the interpretation of facts because every description already involves active interpretation." If we disregard a subjectivism that would have any interpretation as good as any other, the question is why some interpretations are superior to others. Evidently it is not because some "truly reflect reality," and some do not; for there is no uninterpreted reality or facts accessible to us for making comparative judgments. The "pragmatist answer," she says, is "that interpretations do not reflect reality, as in a mirror . . . they are means of utilizing reality congruent with our interests." When interpretations "literally work," she adds, "this success" allows us to say "they worked because they grasped reality." But this "answer" does nothing to explain why some interpretations "work" more successfully than others. (It is here that some notion of truth is indispensable.) The explanation cannot be that some are more successful "because they grasped reality," since 'grasped reality' means here to be more successful. Nor can one rightly say that some are more successful because they satisfy more interests, for that is redundant.

Could there be, after all, some astute truth to the mirror metaphor so distained of late? One grants, of course, that to take the metaphor literally is to lose metaphor altogether.

I shall pursue the above observations in more detail in commenting on both of these able papers.

H. S. THAYER

City University of New York/City College

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

MEANING AND VALUE*

The big issue in metaethics has always been moral realism. But what issue is that? We cannot untendentiously define moral realism as the doctrine that there are moral *facts*, or that some moral sentences state *truths*, or that there are moral *properties* our actions instantiate. For there are pleonastic uses of 'fact', 'true', and 'property' which would render the doctrine acceptable to anyone prepared to make moral judgments. Will you say that Hitler was evil? Then, in the pleonastic senses of 'fact', 'true', and 'property', you will say that it is a fact that Hitler was evil, that it is true that Hitler was evil, and that Hitler had the property of being evil. No clearheaded emotivist ever intended to deny the propriety of such pleonastic restatements of moral judgments.

Moral realism is best defined in terms of a semantic thesis called *cognitivism*. This holds that moral sentences such as 'Eating animals is wrong' are not relevantly different semantically, in terms of the kind of meaning they have, from paradigm fact-stating sentences such as 'Eating animals is a source of protein'. Moral realism may then be defined as the thesis that cognitivism is true *and* there are true moral statements. So defined, the question of moral realism reduces to two other questions: Is cognitivism true? If so, are there true moral statements?

I. COGNITIVISM

To answer the question about cognitivism we need a theory of the semantic features of paradigm fact-stating sentences which help to bestow that status on them; we need to know, in other words, in what *cognitive meaning* consists. A sentence has cognitive meaning, I propose, if it can be used to assert something that can be *believed*, in the strictest, most literal sense of the verb 'to believe'. Belief is the backbone of cognitive meaning.

So consider the two sentences

(1) Lester believes that eating animals is wrong.

and

(2) Lester believes that eating animals is a source of protein.

Cognitivism, on my proposal, is true if (1) can be true when 'believes' is taken to have exactly the same sense it has in (2); for then the sense

* To be presented in an APA symposium on Moral Cognitivism and Realism, December 28. Nicholas Sturgeon and David B. Wong will comment; see this JOURNAL, this issue, 615–6 and 617–8, respectively, for their comments.

0022-362X/90/8711/602–614 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

in which 'Eating animals is wrong' can be used to say that eating animals is wrong will not differ at all from the sense in which 'Eating animals is a source of protein' can be used to say that eating animals is a source of protein.

I am prepared to argue that cognitivism is true, and if I were to make my argument, I would argue for the following three theses, which together entail cognitivism.

The first is the *propositional theory of believing,* which holds, nearly enough, that believing is a relation to referents of 'that' clauses, and that these referents are propositions. To hold this theory is merely to take first appearances seriously. For (2) looks as though it is saying that Lester stands in the belief relation to something referred to by the singular term 'that eating animals is a source of protein'. After all, we may infer 'Lester believes something' from (2), and from it together with 'That eating animals is a source of protein is well confirmed', we may infer 'Lester believes something that is well confirmed'. But what is this thing, *that eating animals is a source of protein,* which is the referent of the 'that' clause singular term? Evidently, it is abstract, in that it has no spatiotemporal location; it is mind- and language-independent, in that it exists and has properties in possible worlds in which there are neither thinkers nor speakers; and it has essentially the truth condition it has, in that, that eating animals is a source of protein is true in any possible world provided that in that world eating animals is a source of protein. So it would seem that referents of 'that' clauses, and a fortiori the objects in the range of the belief relation, are *propositions:* abstract, mind- and language-independent objects that have essentially the truth conditions they have.[1]

The second thesis I call the *pleonastic conception of propositions,* and it is perhaps best introduced via its mate, the pleonastic conception of *properties,* which will be needed soon enough in its own right. Earlier I alluded to the pleonastic sense of 'property' that allows one to move back and forth between 'Albert is humble' and 'Albert has the property of being humble', and even to move from there to 'There is some property that Albert has'. Well, to a first approximation, the pleonastic conception of properties holds that there is no more to the existence of any property than is required to account for the truth of these sentences. To borrow Paul Boghossian's[2] apt way

[1] Here (and elsewhere in this paper) I depart a little from the letter, if not the spirit, of my *Remnants of Meaning* (Cambridge: MIT, 1987).

[2] "Review of Colin McGinn's *Wittgenstein on Meaning,*" *The Philosophical Review,* XCVIII, 1 (1989): 83–92.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of putting it, there is nothing more to being a property than having the *grammar* of a property; properties, to mix metaphors, are merely the shadows of predicates.[3]

Sans metaphor, the pleonastic conception of properties has two parts, the second of which I shall postpone for the next section. The presently relevant part is that, subject to certain qualifications, the singular term 'the property of being *F*', a pleonastic gift of the predicate '*F*', is not liable to reference failure. If '*F*' is a meaningful predicate phrase, then we know that 'the property of being *F*' refers to the property of being *F*. On the pleonastic conception of properties, one cannot make invidious comparisons among grammatical singular terms that ostensibly refer to properties. In whatever sense the property of being a dog is the referent of 'the property of being a dog', so the property of being wrong is the referent of 'the property of being wrong', and both properties are as easily obtained as the expressions that refer to them. The qualifications alluded to are of two sorts. One concerns expressions like 'the property of being married to him', when 'him' fails to refer, and another, which I shall touch on again later, concerns predicates such as 'is a property that doesn't instantiate itself', where respect for the law of noncontradiction appears to preclude our allowing that 'the property of being non-self-instantiating' has reference.

The pleonastic conception of propositions, being *mutatis mutandis* the same as the pleonastic conception of properties, also has two parts, but, again, only the first is relevant to the argument for cognitivism. This holds that, subject to qualifications that need not concern us, the singular term 'that *S*', a pleonastic gift of the sentence '*S*', is not liable to reference failure. If '*S*' is a meaningful sentence in the indicative mood, then 'that *S*' refers to that *S*—i.e., the proposition that *S*. Consequently, if '*S*' is a meaningful indicative sentence, then the 'that' clause in

Lester believes that *S*.

cannot fail to refer to the proposition that *S*. On the pleonastic conception of propositions, one cannot make invidious comparisons among 'that' clauses. In whatever sense the proposition that eating animals is a source of protein is the referent of 'that eating animals is a source of protein', so the proposition that eating animals is wrong is the referent of 'that eating animals is wrong', and both propositions are as easily obtained as the expressions that refer to them.

[3] The metaphor is from D. M. Armstrong, *Universals* (Boulder: Westview, 1989), p. 78.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The final thesis needed to derive cognitivism is that moral propositions can be believed, i.e., the belief relation expressed in (2), which relates Lester to the proposition that eating animals is a source of protein, is also capable of relating him to the proposition that eating animals is wrong.

These three theses entail that (1) can be true when 'believes' has the same sense it has in (2), and that therefore cognitivism is true. So it is thus, in other words, that cognitivism is established: a sentence has cognitive meaning if it means something that can be believed; 'Eating animals is wrong' means that eating animals is wrong; and that eating animals is wrong is something that can be believed.

It remains to argue for these three theses; but since that goes beyond present limits, I must take their truth as a working hypothesis for the remainder of this paper.

II. COGNITIVISM AND MORAL ANTIREALISM: A THEORIST'S CHALLENGE

Given cognitivism, moral realism is true if there are true moral beliefs. And it may seem hard for there not to be. The cognitivist of the preceding section who would deny moral realism must hold that while there are moral properties, none can be instantiated, that while there are moral propositions, none can be true. No wonder those persuaded by the arguments against moral realism typically take those arguments to show that there can be no moral properties. Such an antirealist would deny that (1) can be true when 'believes' in (1) has the sense it has in (2), and her argument for this would run the following course.

If 'believes' in (1) has the same sense it has in (2), then the 'that' clause in (1) is a semantically complex singular term that refers to the proposition that eating animals is wrong. Now, the reference of a 'that' clause is determined by its syntax and the references of its semantically relevant parts, and the reference of 'wrong', if (1)'s 'that' clause refers to a proposition, must, like the reference of any predicate in a 'that' clause, at least include a reference to a property; otherwise, it is impossible to see how to account for the proposition's truth conditions. And if 'wrong' in (1) refers to a property, then, of course, that property is the property of being wrong. But, the argument continues, no such property exists: familiar arguments are available to show that wrongness cannot be identified with any property that is intrinsically specifiable in nonmoral terms, and familiar arguments are available to urge that wrongness cannot be an irreducibly moral property. It then follows that 'wrong' in (1) cannot refer to a property, that the 'that' clause in (1) cannot refer to a proposition, and that (1) cannot be true when 'believes' is read as having its primary sense. At this point, our moral antirealist would

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

have the option of a Mackie-type error theory, which sees the 'that' clause in (1) as necessarily suffering from reference failure, or of noncognitivism, which sees (1) as having a reading in which 'believes' does not have the same meaning it has in (2), a reading which denies that the 'that' clause in (1) purports to have objectual reference and views the sentence as ascribing to Lester a complex *conative* state whose propositional object, if any, is *not* the proposition that eating animals is wrong.

Yet this way of being a moral antirealist is refuted if, as I have suggested, the pleonastic conception of properties and propositions is correct. At the same time, in establishing cognitivism we have hardly established moral realism. How could we have, when the case for cognitivism does not even address the formidable considerations that have for so long been thought to militate against moral realism? Evidently, there is room in logical space for a cognitivist to be persuaded by the case against moral realism. The note of caution is, however, deliberate: the *cognitivist moral antirealist* owes an explanation of *how moral properties, though they exist all right, can fail to be instantiated.* I am a cognitivist tempted by moral antirealism, and it is this explanatory debt I wish to address.

There are, of course, two questions. Why think that moral realism fails? How can the cognitivist moral antirealist account for the inability of moral properties to be instantiated? I must emphasize that my present concern is only the second question. We are to imagine a theorist who is inclined to be persuaded both by cognitivism and by what she takes to be the best case for moral antirealism. Yet she recognizes, and is troubled by, what she perceives as a tension between cognitivism and moral antirealism—namely, the need to make sense of the idea that, while there are moral properties, they cannot be instantiated. It is this tension which I shall try to resolve.

III. HOW TO BE A COGNITIVIST MORAL ANTIREALIST

We are to assume that there are moral properties but that they cannot be instantiated. Then the question is: Why not? Now, there is nothing mysterious about there being properties that cannot be instantiated: for relevant senses of 'cannot', that is the familiar fate of the property of being a round square and the property of being a three-headed human. In these cases the properties cannot be instantiated because they have instantiation conditions that cannot be satisfied—they entail a contradiction, say, or their instantiation is physically impossible. But if there are moral properties that cannot be instantiated, the explanation *will not* be like this. It will not be that they cannot be instantiated because they have instantiation conditions with features that prevent them from being satisfied. If moral

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

properties cannot be instantiated, it will be because *they do not have instantiation conditions*. But what does this mean, and why think it is true? Let me set up direct answers to these questions by first discussing three matters that will help to motivate the answers.

The second part of the pleonastic conception of properties. The pleonastic conception of properties has two parts. The first part, already used in the argument for cognitivism, is that properties are mere shadows of predicates, that (nearly enough) every meaningful predicate '*F*' has its nominalization, 'the property of being *F*', which cannot fail of reference. The second part, now to be introduced, is an explanatory-asymmetry principle. Succinctly put, the principle is that we do not account for two predicates' having the same meaning in terms of their expressing the same property; rather, we account for their expressing the same property in terms of their having the same meaning.

The intended contrast is with what a heavy-duty Platonist would say. This theorist of predicates would press an analogy with proper names. The semantic value—the meaning, if you will—of a name is the thing it names, and to establish that an expression is a genuine name we must identify some object and determine that the expression stands in the relevant semantic relation to that object. There can be no question of establishing that an expression is a meaningful name by showing that it has a use in the language whose specification does not itself mention that the expression is relevantly related to the object that bears it. In the same way, this Platonist persists, there can be no question of establishing that an expression is a meaningful predicate otherwise than by identifying some property and establishing that the expression stands in the relevant semantic relation to it. There can be no question of establishing that an expression is a meaningful predicate by showing that it has a use in the language whose specification does not itself mention that the expression is relevantly related to the property that is its semantic value. Thus, establishing that two people mean the same by 'wrong' requires identifying the property of moral wrongness and showing that each person's use of the word is relevantly semantically related to that moral property.

By contrast, the pleonastic conception of properties holds that we establish sameness of meaning by showing a relevant sameness of *use*, where this use is specifiable without mention of an expression's being related to a property. Given a meaning-determining use for '*F*'—a use whose intrinsic specification will itself refer to no property—we are licensed, if the predicate is ours, to form the singular term 'the property of being *F*' and to be assured of its having refer-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ence. Give '*F*' an appropriate use, and you *ipso facto* have *F*ness; it is not that you start with *F*ness and determine a meaning for '*F*' by having it stand in the right relation to *F*ness. It follows that establishing that two people mean the same by 'wrong' is not a matter of getting hold of the property of moral wrongness and determining that 'wrong' in both their idiōlects stands in the relevant semantic relation to that property. Instead, it is a matter of showing that features of their uses of 'wrong' entitle us to a sameness-of-meaning ascription. Once sameness of meaning is settled, we are entitled to this redescription: 'wrong' for each person expresses the same property—to wit, the property of being wrong.

So much for the pleonastic conception of properties. To be sure, there is the further little question whether it is right, but, as that takes us beyond what there is space for, I have already announced the conditional nature of my argument.

The argument from disagreement. My purpose now is not to argue against moral realism but to sketch one well-known argument so that I might exploit its bearing on the resolution I shall presently offer. The starting point of the argument is the apparent possibility of rationally irresoluble moral dispute. For example, the following seems possible. One person believes that eating animals is wrong, another that it is morally permissible. Neither belief is based on faulty reasoning, on inadequate intelligence or imagination, or on any false nonmoral beliefs; and their disagreement is rationally irresoluble in that no new evidence or further process of ratiocination will yield a convergence of belief. Now, suppose that one of the two propositions in dispute—that eating animals is wrong and that it is permissible—is determinately true, the other therefore determinately false. Then we shall have to say that, although both people are equally intelligent, rational, sensitive, and thoughtful, and even though neither is laboring under any relevant misconception, there is no mechanism of rational-belief formation whereby the person with the false belief can rid herself of her false belief and acquire the knowledge that the other enjoys. But, the argument continues, this belies the supposition that there really is a fact of the matter as to which proposition is true.

There being no fact of the matter as to whether eating animals is wrong hardly shows that there are *no* moral truths. The final stage of the argument urges that rationally irresoluble moral disagreement is possible on virtually *any* issue. For us it is "self-evident" that it is wrong to cause someone pain for the fun of it, but it is easy to imagine a Nietzschean character who held this did not apply to superior beings like himself.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

It goes without saying that there are various ways the argument from disagreement might be resisted, but its moral is the following for one who accepts it. Two people may have radically different ultimate moral principles, moral principles that they cannot justify by appeal to more basic moral principles, and there is, over and above this dispute, nothing to determine that one of these sets of principles is the set of *true* principles. At the same time, these ultimate moral principles provide the conditions that determine a person's application of her moral terms, the instantiation conditions she associates with them. Hence, we can make no sense of a person's moral terms' having *correct* instantiation conditions.

Sameness-of-meaning criteria. One possible response to the argument from disagreement is that the two people are not really disagreeing, and this because there is no proposition, that eating animals is wrong, which one person believes and the other disbelieves. Rather, each means a different property by 'wrong', and what is really going on is that one person believes, so to say, that eating animals is wrong_1 while the other person disbelieves that eating animals is wrong_2. Imagine two linguistic communities within English such that 'fish' is true of whales in one community and false of whales in the other. We can then imagine a dispute between members of the two communities who are unaware of their linguistic differences, one shouting 'Whales are fish', the other 'No, they're not'. But this is no real dispute, for the proposition that the one is affirming is not the one the other is denying. So why not say that the case in question of moral dispute is like this? After all, the two parties to the dispute have, by hypothesis, different ultimate moral principles, and hence different criteria of application for their moral terms.

Yet there are important differences, and they account for why we deny sameness of meaning in the 'fish' case and affirm it in the 'wrong' case, notwithstanding that in both cases the parties to the apparent dispute are operating with different instantiation conditions. (And, by the same token, these differences—i.e., the features now to be listed—explain why, when a person undergoes a change in his ultimate moral outlook, we explain this as a change of mind and not as a change in the meanings of his moral terms.)

First, a mild form of internalism seems to be true. It is arguably constitutive of the concept-fixing conceptual role of 'wrong' that one cannot believe that a kind of behavior is wrong without having *some* disposition not to indulge in it. A corollary of this is that two people engaged in a rationally irresoluble moral dispute will still be disagreeing about *how they want the world to be,* even after they appreciate that their having different ultimate moral principles makes their

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

dispute rationally irresoluble. The imagined "fish" dispute will dissolve when the parties to it realize that each means something different by 'fish', and that, by those meanings, each is asserting a true proposition. But the imagined dispute about eating animals will not have a similar resolution, for even if it were agreed that their different principles *defined* different concepts, they would remain in conflict over how they want the world to be.

Second, we must bear in mind the full range of internalist connections; for in addition to the connection between thinking something wrong and having desires against it, there are also connections with a whole battery of moral emotions. If you think you did something wrong, you are liable to feel guilt, shame, or remorse. You may feel moral pride if you did something especially good, and if someone else does something wrong, you may feel moral outrage or resentment.

Third, it seems to belong to the functional architecture of moral concepts that what ultimate moral views a person has may be partly a function of that person's conative nature, of that person's basic and underived views as to the kind of world he cares to inhabit. The possibility of a change of heart leading to a change of ultimate moral views would explain why it is that, while one regards one's ultimate moral principles (e.g., that it is wrong to cause pain for the fun of it) as necessary truths, one does not regard them as analytic. It would also give point to moral debate even when the debate is rationally irresoluble in the sense explained above; for there would remain the possibility of *affectingly* influencing someone in a way that would lead to a change in his moral views.

Fourth, moral notions come in *families*, and even though two people might associate quite different instantiation conditions with their various moral terms, those terms will belong to structurally identical families of notions. Whatever criteria of application two people have for 'wrong', the relation of that term to their uses of 'right', 'ought', 'good', 'bad', 'just', etc., and to the moral emotions associated with those notions will be the same.

Fifth, moral concepts, however interpersonally dissimilar, are inculcated in a similar process of moral training. Reward and punishment are doled out to the child in the course of training her to be, by her parents' lights, a moral person: "That's *bad!* Don't do it!" "That's not *fair*—give your brother half!" "That's very *wrong* and you should feel very *ashamed!*" (To appreciate how effective this training can be, think of the residual feelings of guilt and self-hatred that people are capable of feeling regarding behavior they now consider morally permissible.)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Sixth, there may be a *common point* to our having concepts with the features I am ascribing to moral concepts. To ask about the point of moral concepts is to ask what they do for us, in what tangible way we would be worse off without them. Consider concepts of shape and color. It is clear how we would be worse off without them: whether our basic needs get satisfied depends on what we do, and what we do depends in part on what we can believe; and without concepts of shape and color we simply could not form many of the beliefs that now enable us to satisfy basic desires; our ability to survive would be greatly diminished if we could not form beliefs involving those concepts. But it is arguable that this sort of story cannot be told about moral concepts. Naturally, I hesitate to speak of *the* point of anything, but putting that qualm aside, I propose that the point of moral concepts is that they help to get people to behave in ways we call "moral."

This returns us to the main order of business. The challenge to the would-be cognitivist moral antirealist is to explain why moral properties, whose existence the theorist affirms, cannot be instantiated. My answer was that moral properties lack instantiation conditions, and this invited me to clarify what it meant and to explain why it is true. Let me turn to this, now that relevant preliminaries have been dispatched.

I take it that nothing can instantiate a moral property unless it has nonmoral properties on which the moral property supervenes. If an act is wrong, then it is wrong in every possible world that is nonmorally indistinguishable from the actual world. I stipulate that a moral property's instantiation conditions are just those nonmoral properties, if any, on which the moral property supervenes.

My explanation of why moral properties fail to have instantiation conditions is as follows. Typically, we judge two people to mean the same thing by a predicate only if they use, or are disposed to defer to, the same criteria for applying the term, but this is not so in the case of moral predicates like 'wrong'. The "criteria" of application a person associates with 'wrong' are determined by her ultimate views as to what is wrong, and there may be great interpersonal differences among these views, these criteria of application. At the same time, important interpersonal similarities of use of the kind just reviewed account for our judging two people to mean the same by 'wrong', notwithstanding the different criteria of application they attach to the term. Here there is nothing to distinguish any one person's criteria as the *correct* criteria of application; this is the result of there being sameness of meaning in the absence of conditions of application that are criterial for everyone. Enter now the pleonastic con-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ception of properties, which, together with the pertinent facts, implies that two people mean the same property of moral wrongness by their uses of 'wrong', however much their ultimate moral views may differ. Now the instantiation conditions of a property are just the correct criteria of application of the predicate that expresses it, but since for moral terms there are no such criteria, moral properties—the mere shadows of moral predicates—have no instantiation conditions. If we thought of moral terms' getting their meaning by coming to express antecedently existing moral properties, then it would be a big mystery why those properties could not have determinate application. What could account for their being such ontological mutants? But on the pleonastic conception, the property expressed by a predicate supervenes on, and has its nature determined by, the features that fix the predicate's interpersonal sense; and we have seen how sameness-of-sense criteria can ascribe sameness of sense in the absence of interpersonal criteria of application. It further follows that moral propositions, the propositions referred to by 'that' clauses containing moral terms, have not the wherewithal to be true. In this way, we see how it is possible for cognitivism and moral antirealism both to be true.

IV. SUMMARY, MARGINAL SOFTENING, AND UPSHOT

The position I am tempted by is the conjunction of cognitivism and moral antirealism. I have limned part of my case for cognitivism but have done even less for the second conjunct. This paper's primary task has been quite limited: to give the most plausible story of how cognitivism and moral antirealism might both be true; to explain, that is, how it might be that, while there are moral properties, they cannot be instantiated, while there are moral propositions, they cannot be true.

I have been concerned with an extreme position, one that allows no moral truths, but I hope it is clear how it could be softened without loss of interest and without giving up its key ingredients. Perhaps it will be agreed that there may be huge interpersonal differences among ultimate moral principles but insisted that there are limits: no one who has our moral concepts can deny that it is wrong to torture people for smoking in public places. But little need change even if this were conceded. For while this would give moral properties some instantiation conditions, it would still leave a great deal that matters to us uninstantiated by moral properties. Morality, the sum of moral truths, would be the feeblest of guides.

Of the many possible objections to what I have advanced, there is one that I would like to anticipate by sort of accepting. I have argued that our practice of individuating senses, and thus of concepts,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

allows us to ascribe sameness of meaning in the absence of interpersonal criteria of application, thus giving rise to properties that have no instantiation conditions. Yet these are surely *defective* practices, so should we not question that they are ours? They are defective practices all right, giving rise to defective properties. But I do not think we should question that they are our practices. Perhaps an analogy will help. Earlier I said that our practice governing the use of 'property' was such that if F is any predicate, then the sentence

a is F.

may be pleonastically rewritten as

a has the property of being F.

But I hastened to add a qualification: logically reflective speakers will not want to restate

Doghood is non-self-instantiating.

(i.e., doghood, not being a dog, is a property that is not instantiated by itself) as

Doghood has the property of being non-self-instantiating.

for if we allow that there is a property of being non-self-instantiating, then we shall be landed with a property that is self-instantiating iff it is not self-instantiating, and that is a contradiction.

This seems to show that our ordinary conceptual and linguistic practices can lead to absurdity. But I do not think it can seriously be denied that it *is* our practice to license the restatement of 'a is F' as 'a has the property of being F', just as it is our practice to license the restatement of 'a is F' as 'It's a fact that a is F'. It is not as if we really employ subtle criteria—criteria perhaps derived from the true inner nature of properties—that allow us to make the move in some cases but not in others; it is rather that we have inherited defective but concept-fixing practices, as if we were taught a game that typically works well but has a defining rule that leads under certain applications to absurdity.[4] Similarly, I would suggest, our entrenched, albeit

[4] Cf. Charles Chihara, "The Semantic Paradoxes: A Diagnostic Investigation," *The Philosophical Review*, LXXXVIII, 4 (1979): 590–618. I should add that it is not obvious that one must deny that there is a property of being non-self-instantiating. Paul Horwich has proposed (in conversation) that it may be possible to deny the truth of

The predicate 'is non-self-instantiating' is true of a thing iff it is non-self-instantiating.

Using pleonastic restatement, this would then allow us to deny the truth of

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

defective, practices lead us to accept properties having no instantiation conditions. We may decry those practices, but they are ours, and to reject them would be to revise our concepts of meaning, property, and proposition. I believe it to be a merit of the deflationary pleonastic conception of properties that it allows us to see how defective linguistic practices can determine defective properties, properties that lack the wherewithal to be instantiated.

What, finally, should be one's stance toward moral concepts and moral language now that we have seen the light? If one believes a single moral proposition, then one is a moral realist; for moral realism is true if there are true moral propositions, and one—or at any rate one who has the concept of truth—cannot believe a proposition without believing that it is true. Since many people have moral beliefs, it follows that they are in error. In this regard, my position is similar to J. L. Mackie's[5] well-known "error theory," although mine differs from his in the argumentation and theoretical apparatus brought to bear and in the fact that my view does not deny the existence of moral properties.

Now, an error theorist need not renounce the *use* of moral notions, for the theorist may recommend the *instrumental* use of moral talk, even if she relegates the benefit of moral concepts to the status of, as it were, instruments of propaganda or subliminal advertising. But even if one decides to live without full-blooded moral notions, this decision would leave unaffected one's desire to live in what before one's decision one would have called the morally best world.

STEPHEN SCHIFFER
City University of New York/Graduate Center

A thing instantiates the property of being non-self-instantiating iff it is non-self-instantiating.

thereby enabling us to allow the existence of the property without fear of contradiction. The problem is a hairy one, and it is not clear what to say, although the very nature of the problem strikes me as offering oblique support for the pleonastic conception of properties, the view that there is no more to the existence of properties than is licensed by our hypostatizing grammatical transformations.

[5] *Ethics: Inventing Right and Wrong* (Harmondsworth: Penguin, 1977).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

SCHIFFER ON MEANING AND VALUE*

Stephen Schiffer's question is, as he emphasizes, a conditional one: if you found the standard arguments against moral realism compelling, but were persuaded by a "deflationary, pleonastic conception of properties" into accepting moral properties, would you have room left for a coherent, motivated position? As instantiated on me, this question is not only doubly conditional but doubly counterfactual. I am far from convinced by the standard arguments against moral realism; and although I believe that there are properties, I would deny that properties, or at any rate genuine, full-fledged properties, exist corresponding to all or even most meaningful linguistic predicates. Moreover, I see these two issues, concerning the defensibility of moral realism and the relation of predicates to properties, as connected in interesting ways, of which one will be the focus of my commentary. For it seems that, in general, the effect of adopting a view of properties like mine, rather than Schiffer's purely pleonastic conception, must be to make moral realism (which holds, at a minimum, that moral properties exist, and are instantiated) *harder* to defend. This is, I think, as it should be. Despite my sympathies for moral realism, in fact, and despite Schiffer's rejection of it, my main complaint about his paper is that he makes the doctrine too easy to accept, and too hard to avoid. To his conditional question, of whether there is room for a position that accepts moral properties but rejects moral realism, he argues that the answer is "yes." What he shows, I think, is that the answer is "just barely" and at a very high cost.

Schiffer describes the position he is exploring as the conjunction of cognitivism with antirealism. As cognitivism is usually understood, that is already J. L. Mackie's[1] position; for Mackie takes moral statements to express beliefs capable of truth values but denies that, as ordinarily understood, any of them are true. Schiffer builds into cognitivism a commitment to moral properties, which Mackie would reject, but recognizes that, like Mackie's, his must be an error theory. The two positions differ strikingly, however, about just who is in

* Abstract of a paper to be presented in an APA symposium on Moral Cognitivism and Realism, December 28, commenting on a paper by Stephen Schiffer; David B. Wong will also comment. See this JOURNAL, this issue, 602–614 and 617–8, respectively, for their contributions.

[1] *Ethics: Inventing Right and Wrong* (Harmondsworth: Penguin, 1977), ch. 1.

0022-362X/90/8711/615–6 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

error. Mackie would convict most of us in our nonphilosophical moments, but not philosophers self-consciously in the subjectivist, conventionalist tradition he praises, and certainly not noncognitivists (though he accuses them of a different mistake). Schiffer, by contrast, can exempt only those who abandon moral language altogether, or else are careful to use it only in an "instrumental" fashion, of which he offers little explanation. (Will he say, "Torture is wrong, but I don't believe that it is, and it's not true that it is"?) Anyone who believes a single moral proposition is, on Schiffer's account, at least implicitly a moral realist. This appears to misclassify philosophers who hold moral beliefs but assign them deliberately deflationary, antirealist truth conditions. More strikingly, it surely misclassifies those "clear headed" noncognitivists who, as Schiffer notes, will accept the equation of 'Hitler was evil' with 'It's true that Hitler was evil', and are happy to say, and think, that Hitler was evil. As a beleaguered moral realist, I suppose I should welcome allies where I find them, but conversion of C. L. Stevenson, R. M. Hare, and Simon Blackburn by this sort of taxonomic gerrymandering seems the wrong sort of victory. I shall argue that moral realism is a more interesting position than this, harder to fall into by misadventure; and I shall explore how, given his views, Schiffer might avoid it without having to paint himself into such a tight corner.

NICHOLAS L. STURGEON

Cornell University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

A RELATIVIST ALTERNATIVE TO ANTIREALISM*

Stephen Schiffer's cognitivist antirealism is appealing in its consistency with the reality of both moral belief and irresoluble moral disagreement. I shall argue for a relativist position with the same advantages, but which is better at discriminating between differences and similarities in the meaning of moral terms, and which need not presuppose the extreme view of moral disagreement that Schiffer's presupposes.

A key claim in his argument is that people can mean the same thing by a moral term while having different criteria for applying the term. He grants that we typically judge people to mean the same thing by a predicate only if they use, or are disposed to defer to, the same criteria for applying the term. He posits sameness of meaning in the moral case because there are pragmatic similarities of use among people who have different criteria for applying a term such as 'wrong'. Because 'wrong' means the same thing on this level, argues Schiffer, it refers to a single moral property, but the fact that there are no uniform criteria for its application means that the property fails to have instantiation conditions.

The pragmatic similarities of use establish that people can talk on some level about the same subject and can morally disagree with each other, even if they do not disagree as to whether common criteria of application have been satisfied. They may conflict in how they want the world to be, as Schiffer points out. But this level at which 'wrong' means the same thing for people with different criteria of application is not the level at which sameness of properties is established. The appropriate conclusion, I shall argue, is not that moral properties fail to have instantiation conditions, but that these properties vary across communities of moral-language users. The relativist alternative allows such communities to employ different criteria for applying moral terms and thus to be referring to different properties, while pragmatically disagreeing over how the world is to be.

One problem for this sort of relativism is the fact that, in moral discourse, disagreement is typically construed not just as pragmatic conflict, but as disagreement over when moral terms are correctly

* Abstract of a paper to be presented in an APA symposium on Moral Cognitivism and Realism, December 28, commenting on a paper by Stephen Schiffer; Nicholas Sturgeon will also comment. See this JOURNAL, this issue, 602–614 and 615–6, respectively, for their contributions.

0022-362X/90/8711/617–8
© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

applied. The relativist can explain this construal by pointing out that fundamental differences among sets of criteria for applying a moral term are accompanied by broad similarities and overlapping among these sets. The similarities and overlapping, together with pragmatic similarities of use, obscure the differences, and make it easy to conflate pragmatic disagreement with disagreement over the correct application of moral terms.

My version of relativism denies Schiffer's postulate that irresoluble moral disagreement permeates all moral discourse, and it also denies that such disagreement is anomalous. There is a tendency in moral philosophy to consider only the extreme possibilities, but our criteria for applying moral terms are diverse. Some may be the subject of broad interpersonal and intercultural agreement, while others may be the source of the kind of irresoluble disagreement Schiffer mentions.

Finally, I shall argue that concept-fixing practices that do not yield uniform criteria of application are "defective," only on the assumption that the function of morality is to identify moral properties that in principle can be universally acknowledged and then used as guides to action. On an alternative conception, morality functions to regulate conflicts of interest both between and within individuals. Irresoluble disagreements stem from fundamental differences between ways of meeting this regulative need. Such differences can be deeply disturbing, and may frustrate some deeply rooted human needs, but are not necessarily signs of defect.

DAVID B. WONG

Brandeis University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

A PROBLEM IN THE FOUNDATIONS OF SET THEORY*

The problem I want to consider has many baffling aspects, the most fundamental of which is the distinct possibility that it is actually a pseudo problem. My plan is to isolate one perspective from among those which seem least congenial to the problem's legitimacy and to explore the view from there. Be forewarned that I shall reach no firm conclusions; this is a purely exploratory expedition.

The problem arises from a question independent of the standard axioms of set theory: Are certain sets of real numbers, the so-called Σ^1_2 sets, Lebesgue-measurable? This question, unlike the more familiar continuum hypothesis, does not depend on the notion of "all sets of reals," and those few sets of reals it *is* concerned with are relatively simple, defined in terms of basic, geometrically concrete operations. It arose naturally in the course of a mathematically straightforward inquiry, not in the midst of a bold conceptual remodeling like Cantor's. In fact, it arose in the context of work intended to counteract the nonconstructive tendencies of modern set theory.[1]

Also, more is known about this problem than about the continuum hypothesis. Godel's work on constructibility implies that, if the axiom of constructibility ($V = L$) is true, then there is a nonmeasurable Σ^1_2 set. In contrast, early work of Solovay on large cardinals shows that, if there is a measurable cardinal (MC), then all Σ^1_2 sets are measurable. Thus, if the question of the measurability of Σ^1_2 sets is a genuine problem, one way of solving it would be to find a principled way of choosing between $V = L$ and MC.

This "problem in the foundations of set theory" is actually three interdependent problems: (1) Is "are there nonmeasurable Σ^1_2 sets?" a genuine problem or a pseudo problem? (2) What counts as evidence for or against the claim that there are nonmeasurable Σ^1_2 sets? (3) Are there nonmeasurable Σ^1_2 sets? My main concern here is the first of these.

One who takes an independent statement as a genuine open problem is standardly accused of being a Platonist. As many consider Platonism a nonstarter these days, I have tried in recent years to

*To be presented in an APA symposium on Realism and the Philosophy of Mathematics, December 29. Hartry Field will comment. My thanks to J. Barwise, J. Etchemendy, J. Kvanvig, and A. Nelson and the members of his Winter 1990 seminar, and to the University of California/Irvine.

[1] See my *Realism in Mathematics* (New York: Oxford, 1990), ch. 4.

0022-362X/90/8711/619–628 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

revive it. But to tie the legitimacy of this problem too closely to Platonism would distort its true sources. In fact, the problem is also genuine for (at least some) structuralists, modalists, and even nominalists.[2]

There is, however, one principle that unites all these thinkers: scientific realism. Even as mathematics is reinterpreted, rearranged, uninterpreted, the natural sciences remain unquestioned. But notice: in the philosophy of science itself, as opposed to the philosophy of mathematics, recent history has not been so kind to scientific realism. It has been attacked in various ways by numerous writers, some of whom have taken the trouble to pay close attention to the actual inferential practices of scientists themselves.[3] The potential impact of these critiques on the underpinnings of contemporary philosophy of mathematics cannot be ignored. My goal here is to investigate the consequences for our particular problem.

In my exploration of scientific antirealism, I shall take along one familiar compass: naturalism. By 'naturalism' I understand the simple Quinean idea that we should not expect to view science from some extrascientific perspective, that we should instead judge science by the standards of science itself. As with other Quinean insights, naturalism is expressed perhaps most forcefully by a departed and much-mourned Putnam-stage[4] who once wrote:

> . . . it is silly to agree that a reason for believing that *p* warrants accepting *p* in all scientific circumstances, and then to add, 'but even so it is not *good enough*'. Such a judgement could only be made if one accepted a trans-scientific method as superior to the scientific method; but this philosopher, at least, has no interest in doing *that* (*ibid.*, p. 356).

Notice that naturalism could cut the other way as well: if a reason for believing *p does not* warrant accepting *p* in all scientific circumstances, then it will not do to insist "but even so it *is* good enough."

Some identify scientific realism with an extrascientific correspondence theory of truth, but this will not work in a naturalistic context. I prefer an old-fashioned ontological characterization: the scientific realist believes there are genes, molecules, electrons, etc., and resists operationalist, instrumentalist, and various other reductionist efforts

[2] See *Realism in Mathematics*, ch. 5.

[3] Examples: Nancy Cartwright, *How the Laws of Physics Lie* (New York: Oxford, 1983); Ian Hacking, *Representing and Intervening* (New York: Cambridge, 1983); and Arthur Fine, *The Shaky Game* (Chicago: University Press, 1986).

[4] Hilary Putnam, "Philosophy of Logic," reprinted in his *Mathematics, Matter and Method* (New York: Cambridge, 1979, 2nd ed.), pp. 323–357.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

to reinterpret these claims.[5] The crucial debate centers on the existence of unobservables, "crucial" because it is here that pseudo problems are said to arise. If the theory of electrons, say, is understood as a mere fiction, a useful instrument, then to ask a question undecided by the theory is akin to asking about the length of Hamlet's nose. If, on the other hand, there are electrons, then such a question may well be genuine.

So, why believe in electrons? A standard realist response is that they figure in the best explanation of our experience; thus inference to the best explanation assures us that they do exist. But, the antirealist insists, the fact that a theory explains does not show that it is true.[6] For the antirealist:

> . . . the high level explanations of theoretical science . . . organize, briefly and efficiently, the unwieldy, and perhaps unlearnable, mass of highly detailed knowledge that we have of phenomena. But organizing power has nothing to do with truth.[7]

Here the antirealist is inspired by a picture of the world very different from the realist's. Nancy Cartwright characterizes the difference like this:

> Richard Feynman talks about explaining in physics as fitting phenomena into 'the patterns of nature'. But where are the patterns? Things happen in nature. Often they happen in regular ways when the circumstances are similar . . . there are analogies between what happens in some situations and what happens in others. . . . But there is only what happens, and what we say about it. Nature tends to a wild profusion, which our thinking does not wholly confine. The metaphysical picture [here] is an Aristotelian belief in the richness and variety of the concrete and particular. Things are made to look the same only when we fail to examine them too closely (*ibid.*, p. 19).

The realist sees nature as inherently patterned and sees good explanations as likely to latch onto these patterns. The antirealist sees nature as a chaotic jumble of particular occurrences into which our explanations can sometimes bring a bit of artificial order. From this second point of view, explanations are valuable, but they are not likely to be true.

[5] Paul Horwich ["Three Forms of Realism," *Synthese*, LI (1982): 181–201] calls this "semantic realism." I find it hard to distinguish from Fine's "Natural Ontological Attitude."

[6] Bas van Fraassen, *The Scientific Image* (New York: Oxford, 1980); Cartwright, *op. cit.*; and Richard Miller, *Fact and Method* (Princeton: University Press, 1987).

[7] Cartwright, p. 87.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

The potential repercussions of this line of thought for the status quo in contemporary philosophy of mathematics are immediate and widespread. Much of the current debate is shaped by the assumption that the Quine/Putnam indispensability arguments provide strong grounds for supposing that at least some mathematical claims are true. Platonists (like me) and many others (like the aforementioned structuralists and modalists) take the indispensability arguments as fundamental. Even the nominalists lavish much of their energy on efforts to show that various parts of mathematics are in fact dispensable, despite appearances. But the indispensability arguments themselves typically presuppose a scientific realism based on inference to the best explanation: mathematics figures in our best explanation of phenomena; as scientific realists, we take these explanations to be true; therefore, we must believe in the entities that mathematics asserts to exist.

This is the uncongenial philosophical perspective I had in mind at the outset: the antirealist perspective that questions inference to the best explanation. My plan is to adopt it and to explore its consequences for our problem about $\sum^1_2$ sets.

Of the antirealists, van Fraassen draws the most sweeping conclusions from the rejection of inference to the best explanation. He holds that we have good reason to believe in what we observe, but that it is always reasonable to refrain from believing our best explanatory theories of unobservables. Instead we "accept" these theories, which is to say that we hold them to be "empirically adequate," which is to hold that what they say about observables is true. In other words, we do not infer to the truth of our best explanation, but (at best) to its empirical adequacy.

Clearly, the naturalist must reject this form of empiricism. It is by definition unnaturalistic to admit that electrons pass all relevant scientific tests for existence and still insist we can reasonably refrain from belief. Here the extrascientific standard being applied is something like: believe only in what is accessible to unaided acts of human perception. This principle is not part of scientific practice, as numerous case studies show. Van Fraassen would agree that, from within our scientific theory, 'electron' corresponds to something in the actual world while 'flying horse' does not, but he recommends that we "bracket" this ontological commitment and step back to an extratheoretic standpoint to evaluate our "epistemic attitude" (*op. cit.*, pp. 81–2). But this is just the sort of move the naturalist rejects.

This is not to say that van Fraassen's acceptance/belief distinction is of no use to the naturalist. To see this, notice that, in fact, infer-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

ence to the best explanation has a mixed record of success. Berzelius's reasoning from the spacial arrangement of atoms in molecules to the discovery of isomers assumed the truth of the atomic hypothesis; on the other hand, Rayleigh was led astray in his theory of radiant heat by his literal belief in the (current) best explanation, that is, in ether as a universal medium.[8] If inference to the best explanation is not the dependable rule realists often take it for, we should expect scientists to know this. And lo, the case studies suggest that they do, that in practice scientists do distinguish between acceptance[9] and belief, though the line does not fall where van Fraassen would have it. This is just the sort of thing a naturalist is duty-bound to take seriously, and I shall.

Consider the case of molecules. Between 1800 and 1860, the molecular hypothesis was used to explain the laws of definite and multiple proportions and of combining volumes and substitution, as well as the existence of isomers. Impressed by this overwhelming explanatory success, scientists agreed to accept the hypothesis as adequate. On its truth, however, there was no agreement. Even as late as 1904, a canonical textbook reads:

> . . . the atomic hypothesis has proved to be an exceedingly useful aid to instruction and investigation, since it greatly facilitates the interpretation and use of the general laws. One must not, however, be led astray by this agreement between picture and reality and combine the two.[10]

Here a sizeable and responsible portion of the scientific community refuses to apply inference to the best explanation. And it seems unlikely that the doubters simply held out for a higher degree of explanatory success; after all, the theory was successful enough to qualify for unanimous acceptance as adequate. For the naturalist, this case runs opposite to van Fraassen's empiricism; this time the realist would have us certify an inference as good enough despite the reluctance of practicing scientists.

The situation changed completely within the next few years, mainly as a result of Perrin's experiments on Brownian motion. The 1908 textbook by the author just quoted reads:

> . . . *we recently came into possession of experimental proofs of the discrete or particulate nature of matter, proofs which the atomic hy-*

[8] Miller, pp. 453–461.

[9] Here 'acceptance' means "belief in adequacy," where a theory is adequate if what it says about real (believed in) things is true.

[10] This and the next two quotations come from Miller, pp. 473–4. See also M. Nye, *Molecular Reality* (New York: American Elsevier, 1972).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

> *pothesis vainly sought for centuries, indeed, for millennia.* . . . the agreement of Brownian motion with the demands of the kinetic hypothesis, which was established by a series of investigators, finally and most completely by J. Perrin, justify even a cautious investigator in speaking of an experimental proof of the atomic nature of extended matter (*ibid.*).

Even the long-time skeptic Poincaré was moved to write that "atoms are no longer a useful fiction . . . The atom of the chemist is now a reality."

In recent writings, several philosophers of science have urged that contrasts like that between the pre-1860 evidence that led to acceptance and the post-1904 evidence that led to belief be seen as a difference in kind rather than degree. From the Cartwright's Aristotelian perspective, there is no reason to expect that good explanations, that is, good organizing devices, will be true to the "richness and variety of the concrete and particular," and Miller suggests, more specifically, that phenomena like the chemical combination ratios Dalton explained could be seen as "regular tendencies standing in need of no [true] explanation" (*op. cit.*, p. 476). Brownian motion, however, is a different matter. Cartwright would say that Perrin inferred to the most likely cause, not to the best explanation; Miller would say the convincingness of Perrin's experiments depends on the "topic-specific truism" that "it is unreasonable to suppose that nonliving things that constantly deviate from their courses in an irregular manner do so without being caused to deviate by something external to them" (*op. cit.*, p. 480); and Hacking would probably hold that Perrin's work on molecules marked the beginning of "our ability to manipulate them using well-understood low-level causal properties" (*op. cit.*, p. 274). Thus, inference to the adequacy of the best explanation in the pre-1860 phase contrasts with more concrete inference to the actual existence of molecules in the post-1904 phase.

Notice that justifications of these latter sorts will support literal belief in the entities in question, but perhaps not in all the properties the previously accepted theory attributes to them.[11] Room is left for part of the theory to be accepted and part to be literally believed. And this distinction has methodological force:

> 'Competing' theoretical treatments . . . are encouraged . . . but only a single causal story is allowed.[12]

[11] This is a major theme of Cartwright and Hacking.
[12] Cartwright, pp. 85–6.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

This is a moderate realism. It is not standard realism; inference to the best explanation as a general rule is rejected, along with an unquestioning belief in the truth of our best explanatory theories. Yet it is not standard antirealism, either, because disbelief in some unobservables is held to be unreasonable in the light of certain kinds of evidence. And finally, this moderate realism lays claim to being a more accurate description of actual scientific practice, a fact of considerable importance to the naturalist.

What does all this mean for set theory? Alas, even this moderate realism calls the status of the indispensability arguments into question. We must resist the impulse to argue for the truth of mathematical claims solely on the grounds that they figure in our best explanation of experience. We must face the possibility that all mathematical parts of scientific theory fall within the range of acceptance rather than belief. In which case, it might well be confused to expect determinate answers to questions beyond the reach of our current assumptions.

We need to know how far literal belief extends. Perhaps, for example, Cartwright's inference to the most likely cause will end up committing her to the literal truth of various straightforwardly mathematical claims. She writes:

> . . . the propositions to which we commit ourselves when we accept a causal explanation are highly detailed causal principles and concrete phenomenological laws, specific to the situation at hand, not the abstract equations of a fundamental theory (*ibid.*, p. 8).

So, for example, in the case of the radiometer, she contrasts literal belief that

> [The] velocity (with which the gas slides over the surface) and the corresponding tangential stress are affected by inequalities of temperature at the surface of the solid . . . (*ibid.*, pp. 8–9).

with mere acceptance of Boltzmann's equation. To connect the qualitative description with the equation, we produce a model:

> A model is a work of fiction. Some properties ascribed to objects in the model will be genuine properties of the objects modelled, but others will be merely properties of convenience . . . [these] are introduced into [the] model . . . to bring the objects modelled into the range of the mathematical theory (*ibid.*, p. 153).

Clearly much of the mathematical content of the theory is not to be literally believed. But the literally true description quoted above

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

 Digitized by Arya Samaj Foundation Chennai and eGangotri

includes a reference to the temperature of a surface, and elsewhere, Cartwright admits that

> Some claims of the theory must be literally descriptive (I think the claims about the mass and charge of an electron are a good example) if the theory is to be brought to bear on phenomena (*ibid.*, p. 78).

Perhaps this is enough to force literal belief at least in the real numbers.

In fact, I do not think this necessarily follows from what Cartwright says. Claims about mass or temperature may well be literally true, but perhaps the identification of these quantities with functions from objects to real numbers occurs only in the process of mathematical modeling. In the course of his nominalistic project, Hartry Field[13] points to purely physical facts about quantities like temperature which make possible the application of real numbers as labels. I doubt that Cartwright need commit herself to any more than these, and perhaps she can get away with even less. For example, it is not clear to me that she need assume, prior to modeling, that masses or temperatures vary continuously. If not, then her literally believed reality may not even include the sorts of continuous phenomena that end up bringing some nominalists into such close comradeship with the Platonist.

These issues are hard to judge, partly because the status of mathematical entities is not explicitly at issue in the writings of the moderate realists, but I suspect a viable case for their reality in this context must be pitched elsewhere than on the use of real numbers as labels. As naturalists, we might begin by asking why modern infinitary mathematics was introduced in the first place. In response, I think it is enough to consider motion. To understand how motion is possible, and to control and predict it, scientists needed two notions—function and continuity—which bring with them the completed infinite and our modern conception of space. The question we are concerned with—are $\sum^1_2$ point sets Lebesgue-measurable?—arises when this theory of space is taken literally. To ask it is to think like a post-Perrinean chemist; given that there are such things—molecules or point sets—it makes sense to inquire into their further properties—the precise spacial arrangement of atoms in a given type of molecule or the measurability of a given type of point set.

Looked at in this way, what we need to know in order to assess the legitimacy of our problem is whether the continuity of space is real or

[13] *Science without Numbers* (Princeton: University Press, 1980).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

an artifact of the mathematical modeling. Untutored intuition (at least mine!) suggests that an argument for literal continuity might begin from the success of the calculus. The criteria of the moderate realists often depends on a sense of what needs explaining and what could count as a mere regularity, and motion seems to me a prime example of something that needs explaining. (How does Zeno's arrow move?) And notice: if space-time is causally efficacious, as some say, the evidence for it might even satisfy one or another of the moderate realists' broadly causal criteria. Of course, philosophers of physics still debate the reality of space-time, but philosophers still debate the reality of molecules! What matters is what the scientists say, and what evidence they cite.

For me at least, this question requires further study. My main goal here is to locate the point at issue for the genuineness of our problem, and I take the continuity of space to be a strong candidate. Caution is in order, however; there are ways of eliminating space-time from physics which would not affect the genuineness of our problem simply because they put other continuous phenomena in its place. For example, John Burgess[14] has shown how space-time can be dispensed with, assuming that "for some (one-dimensional or more) continuum, every point-instant inside it is occupied" (*ibid.*). And Field[15] discusses the elimination of space-time in favor of fields, in which case the field itself will be continuous, at least in parts. So it would be better to formulate the point at issue as the reality of some continuous thing or other. Even continuous phenomena completely unrelated to the theory of space might give substance to our problem.

Finally, let us glance at what hangs on the outcome. If ours is in fact a pseudo problem, we might still wonder whether a theory with nonmeasurable Σ^1_2 sets would be more acceptable than one without. This might be decided on grounds of consequences for parts of our theory that we take literally, or, more likely, on purely pragmatic grounds. There is even the option of using different axioms for different problems, or even for the same problem viewed from different angles.

If, on the other hand, the problem is genuine, our naturalistic search for a solution should begin with a study of actual practice. This I have done in some detail, but the results are inconclusive.[16]

[14] "Synthetic Mechanics Revisited" (unpublished manuscript).
[15] *Realism, Mathematics and Modality* (New York: Blackwell, 1989).
[16] See my "Believing the Axioms," *Journal of Symbolic Logic*, LIII (1988): 481–511, 736–764.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

We find many wonderful theorems and some scattered remarks about what might count as evidence for what, but no clear sense of proper method and no persuasive consensus. What makes our problem a particularly fascinating one for philosophers is exactly what makes it a particularly tricky one for the naturalist, that is, the methodological uncertainty among actual practitioners. To find an answer would be to help *develop* a methodology, not merely to describe one.

This is not a standardly mathematical problem—a proof is not what is called for—but various mathematicians, from Zermelo and Godel to contemporaries like Martin and Jensen, have discussed method. Perhaps this is presumptuous, but to me the current confusion suggests that a real contribution might be made by philosophically trained observers, if not me, then one of you. This is not to advocate an appeal to extrascientific philosophical standards. What is needed is insight into how to proceed, insight that makes sense to practitioners. A purist might insist that a philosopher who offers such insights is acting as a mathematician in doing so, but I do not think this matter of professional affiliation is really worth arguing about. I hope to have shown that there are more important issues at stake.

Of course, it is risky to devote one's energies to a problem that may turn out to be a pseudo problem after all, but some of the greatest mathematicians of the recent past and present have been doing this for decades. Mathematicians do not make the mistake of asking whether groups are commutative; some are, some are not. They do make the mistake, if it is one, of asking whether Σ^1_2 sets are measurable. Understanding how this has come about, whether the problem turns out to be genuine or not, should eventually lead to a better grasp of the nature of mathematics and of its role in science.

PENELOPE MADDY

University of California/Irvine

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

AFRICAN FAMINE: NEW ECONOMIC AND ETHICAL PERSPECTIVES*

It is widely believed that recurrent episodes of famine in Ethiopia, the Sudan, and in other nations of the sub-Saharan African "Sahel region" are caused by severe drought and by the resultant destruction of productive agricultural capabilities in these countries. It is also alleged that these disasters are caused by environmental catastrophes: by "desertification," for example, in which the Sahara desert extends its reach southward at a rapid pace, owing to poor animal husbandry, ill-advised agricultural policies, and general overpopulation.

Our moral response to these tragedies is strongly dictated by our beliefs concerning the causes of these disasters. And it is the main thesis of this paper that these customary understandings of the causes of famine are seriously mistaken. I shall argue that neither natural disasters (like drought or flood), nor environmental collapse, nor even extreme population expansion, "cause" famines in any direct sense. By extrapolating the results of an alternative microeconomic analysis of famine proposed in 1981 by Amartya Sen, I shall suggest more appropriate moral and political patterns of response to these tragedies.

I

The most recent famine in Ethiopia, which began in September 1984, appears to follow a recurring pattern. Scarcely a decade separated the onset of the 1984–86 disaster from a similar five-year dry spell during the years from 1969 to 1974, which resulted in equally severe famines in Ethiopia and in the neighboring countries of the sub-Sahara African Sahel region during the period 1972–74.

These earlier events in Africa during the 1970s were accompanied by a famine of massive proportions in the then newly created nation of Bangladesh. Precisely at that time, many other nations experienced an unprecedented energy crisis, accompanied by widespread economic instability. Many experts concluded at the time that world economic and environmental systems had been strained to the point of collapse, and predicted that the international political situation could not be far behind. In the resulting atmosphere of fear and

*To be presented in an APA symposium on African Famine, December 29. William Aiken will comment; see this JOURNAL, this issue, 642–3.

0022-362X/90/8711/629–641 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

panic, grim forecasts of widespread famine, energy riots, and global ecological disaster were the norm.[1]

Prior to the early 1970s, famine had been viewed as a specific and localized tragedy brought about by a natural disaster: a flood, an earthquake, or a lengthy drought. The most appropriate response was thought to consist (when possible) in massive donations of food, clothing, and medical supplies until the emergency had abated.

In contrast to this older donation strategy focused upon isolated incidents of famine, the crisis atmosphere of the mid-1970s produced two, sharply opposing schools of thought. Advocates from both schools agreed that famine must be understood as a recurring, cyclical, and systematic phenomenon that requires long-range strategies in response. The sharp disagreement came over which strategies were the most appropriate, in light of radically differing views about the underlying cause of famine.

1. *Developmentalists* argue that, apart from the side effects of war,[2] the principal cause of famine is a sudden decline in the available supply of food (the food availability decline theory or FAD). The problem of inadequate food supplies, moreover, must be met by policies aimed at a permanent restructuring of agricultural policies in the nations most severely affected, rather than by emergency donations of food. The agricultural sectors of impoverished nations must be drastically overhauled and modernized through the introduction of more efficient farming and irrigation technology and the latest agricultural techniques.

In accord with developmentalist views, new high-yield, disease-resistant strains of wheat and rice developed by Nobel-laureate agronomist Norman Bourlag were introduced in many developing nations of the Third World, together with irrigation projects for water and hydroelectric power. Improved technologies, such as modern tractors and combines, were introduced along with, in many cases, proposals for the development of new "cash crops" (such as coffee, cotton, flowers, and vegetables for export to Northern Europe) to provide greater diversity and wealth in these primitive economies, and to shield their citizens from the devastating effects of periodic fluctuations in the weather.

These policies were not ultimately successful. The new seeds of Bourlag's "Green Revolution" required massive amounts of expen-

[1] For an account of this period, see my "The Political and Economic Dimensions of Hunger," in *Lifeboat Ethics: The Moral Dilemmas of Hunger*, Lucas and Ogletree, eds. (New York: Harper & Row, 1976), pp. 1–28.

[2] For the purposes of abridgement, I must delete a lengthy discussion of famine as either unintended outcome or deliberate strategy in wartime.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

sive fertilizer, a form of energy that poorer farmers could not afford. The increased dependence upon cash crops diverted the most valuable and productive land and water supplies in many poorer countries from the direct production of indigenous food supplies, and, in addition, subjected the farmers to the vagaries of international economic competition and recession due to massive fluctuations in crop prices and demand on the world market. The sophisticated agricultural machinery—the tractors and irrigation pumps—required expensive and scarce petroleum, putting impoverished Indian and African farmers into a direct (and unwinnable) economic competition for petroleum with affluent automobile owners in North America and Northern Europe. The new farm machinery was often complicated to operate, and there was an absence of spare parts (not to mention expertise) to effect needed maintenance and repairs. A brief period of improvement in the global food supply, following the implementation of developmentalist policies in the late 1960s, was succeeded by even more virulent and widespread famines in the early 1970s.

2. *Neo-Malthusians.* Citing these dismal results, critics of developmentalist policies argued that it was not the inadequate *supply* of food and scarce natural resources (water, energy, land, and fertilizer) that was in imbalance. Rather, it was the *excessive demand* for such resources, brought on by overpopulation, that caused environmental destruction, economic collapse, and famine. If the population of a given region—Ethiopia or the Sahel, for example—grew at a rate in excess of that local environment's ability or "carrying capacity"[3] to feed, house, and clothe the population, then natural disasters (deaths from famine and disease) would follow inevitably. Indeed, developmentalist strategies had the paradoxical effect of worsening, instead of improving, the situation by allowing the local populations to grow even further beyond their natural ecological limitations, putting further destructive stress upon those fragile local environments and hastening their ultimate and irrevocable demise.

These environmentalists or "neo-Malthusians,"[4] led by American

[3] "Carrying capacity" is defined by biologist Garrett Hardin as the maximum human population sustainable with the resources locally available in a regionally-defined environment; "Carrying Capacity as an Ethical Concept," in *Lifeboat Ethics*, pp. 120–134. See also "The Tragedy of the Commons," *Science*, CLXII (December 13, 1968): 1243–1248; and "Living on a Lifeboat," *Bioscience*, XXIV, 10 (October 1974): 561–8.

[4] I distinguished between advocates of donative, development, and "selective allocation" strategies (following the model of medical triage) in "The Political and Economic Dimensions of Hunger" (1976). O'Neill subsequently utilized the label "neo-Malthusians" for advocates of selective allocation strategies; *Faces of*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

biologist Garrett Hardin, took the controversial view that no aid ought to be given to famine stricken nations, unless the recipients took steps to curb their population growth, and to halt the damage to their local environments caused by, e.g., overgrazing in the Sahel, or excessive cutting of timber in Nepal. These population ecologists argued that—since human beings are largely at the mercy of the weather, and since local regions seemed to have brought disaster on their own heads through intemperate population growth, along with destructive and wasteful agricultural policies—the victims of these disasters in the future must simply be left to starve unless and until the "natural ecological balance" was restored.

II

This acrimonious debate obscures serious explanatory deficiencies in both approaches. While the weather is known to be impossible to control and difficult to predict, it has proven equally difficult over the past two decades to do much about local population growth. The hoped for "demographic transition" as the linchpin of developmentalist theory[5]—in which rising affluence would lead to an automatic downturn in family size—cannot be verified even in the historical experience of affluent countries. Changes in economic policies and personal consumption habits among the world's more affluent populations do not translate readily into increased supplies or lower commodity prices for the world's poor. And neither approach seems able to account for the puzzling but persistent fact, in tragedies ranging from the Great Irish Potato famine of 1848 to Bengal in the 1940s and Ethiopia in 1974, that the most severely affected regions simultaneously engage in the export of food.

The developmentalist approach suggests that a famine ought to occur whenever the following takes place:

> (1) A decline occurs in the overall available supply of food, S, below a critical minimum level, S_o, due to drought, insect infestation, flood, or other natural disaster (the FAD theory).
>
> (1′) A sudden sharp decline occurs in S_c, the per capita supply of food, below a critical minimum threshold, S_c^*.

Hunger: An Essay on Poverty, Justice and Development (London: Unwin Hyman, 1986); cf. also her account of this debate in "The Moral Perplexities of Famine and World Hunger," in *Matters of Life and Death*, Tom Regan, ed. (New York: Random House, 1980/1986), pp. 300–8.

[5] The demographic transition theory holds that "an assured food supply [is essential] in reducing [the need for] high birth rates [in developing countries]." See Lester Brown, *In the Human Interest* (New York: Norton, 1974), p. 119; and *World without Borders* (New York: Vintage Books, 1973), pp. 140–1. For a discussion of the history and significance of this theory, see my "The Political and Economic Dimensions of Hunger," pp. 18–20.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(1″) A sudden and sharp fluctuation takes place in the overall food supply, S, or in the per capita food supply, S_c, as compared to preceding years, even though neither reduction falls below some alleged critical threshold, S_o or S_c*.

According to the neo-Malthusian approach, famine occurs when:

(2) There is a sustained period of excessive population growth, P, in excess of local environmental carrying capacity, P_o.

(2′) There occurs a sudden fluctuation in population growth, an "explosion" relative to immediately preceding population figures, even though P does not necessarily exceed P_o.

(2″) An overlarge population practicing some form of anti-environmental agriculture or animal husbandry does serious environmental damage, thereby lowering the threshold value of P_o relative to P for that region.

Surprisingly, however, there is no clear correlation between fluctuations in a nation's overall food supply and the outbreak of famine. In Ethiopia, for example, available records show that there are years in which the overall food supply drops precipitously low or changes suddenly and dramatically from the preceding year, and yet no famine occurs then or immediately following. By contrast, there are years when the food supply is stable or even improving, and seems more than adequate to feed the existing population, and yet a famine occurs (as was the case both in 1972–74 and 1984–86).

With respect to neo-Malthusian claims, the same anomaly occurs. There are years during which the population of a country like Ethiopia seems to grow at an alarming rate or to change suddenly relative to preceding years, but no famine occurs then or immediately afterward. There are other periods when the population is not excessive and the growth rate acceptable, and per capita food supply more than adequate, and a famine nonetheless breaks out. The same absence of correlation can be noted with regard to alleged ongoing environmental damage being done in sub-Saharan Africa and Ethiopia, as well as serious periodic natural disasters—insect plagues, dust storms, and drought—that may be magnified in the wake of such damage. In short, then, the available data falsify both theories. Neither the ensemble of developmentalist hypotheses nor the set of neo-Malthusian hypotheses describes even necessary conditions, let alone sufficient conditions, for the outbreak of famine.

According to United States Department of Agriculture statistics,[6]

[6] *Sub Saharan Africa: Outlook and Situation Report* (Washington, D.C.: USDA Economic Research Service), esp. the July 1984 and July 1985 editions.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

for example, Ethiopia began experiencing its most recent severe drought in 1979. The sharpest drop in agricultural and food production was not recorded in 1984, when the most recent famine emerged. Rather, there was a sharp drop in agricultural productivity during the period 1979–80 (the first year of the drought). Overall food production was relatively stable thereafter from 1980 through 1984, at levels above the annual average recorded during an extremely productive agricultural period from 1969 to 1971 (to which all subsequent production statistics are normalized).[7] That is, during a period when there was a precipitous change in the overall and the relative food supply, no famine was recorded. Vice versa, during a period of relative stability and normal production and supply, a terrible famine attracting international attention did occur.

Per capita food availability statistics during this same period, however, are more ominous, reflecting an annual 2.5 percent growth rate in the population of Ethiopia during the years 1979 through 1984.[8] Even with a steady, sharp increase in the total quantity of food imports into Ethiopia during this same period, the per capita statistics reveal a steady decrease in the quantity of food available per person for the nation as a whole. These same statistics also suggest, however, that the average quantity of food available to each person in 1983 and 1984 was still in excess of the amount available between 1976 and 1978 when no famine was recorded.

Of particular note, the figures on food availability and population growth, published just before the massive and widely publicized outbreak of famine in September 1984, do not suggest that a crisis was imminent as a result of the ongoing African drought. Indeed, one analyst reviewing the data in July 1984, only a few weeks before the onset of the famine crisis, commented

> During the 1984 spring planting, heavy rains have ended Somalia's drought while Ethiopia has received sufficient rainfall in the central and western areas. . . . Ethiopia's grain output is likely to expand and, given current stocks, imports could decline.[9]

Notice in particular this report concerning the "ending of the drought" in Somalia, and the adequate rainfall in Ethiopia's central regions, issued in July. During the Fall of 1984, however, newspapers and television stations the world over were carrying reports of drought and the outbreak of famine as early as August and Sep-

[7] "Indices of Agricultural and Food Production in Sub-Saharan Africa," *Sub Saharan Africa: Outlook and Situation Report*, RS 84-10 (July 1984), p. 31.
[8] "Basic Indicators of Food-sector Performance," *ibid.*, p. 23.
[9] Susan Buchanan, USDA publication RS 84-10, p. 11.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tember in Ethiopia's northernmost provinces of Eritrea and Tigre, and in the central province of Wollo.[10] In Wollo—where rain was reported as having fallen the preceding spring—the famine was most severe, as it was also in 1974. The news accounts made no mention of the rainfall reported three months earlier, focusing instead upon "the drought" as the cause of the famine outbreak.

III

In *Poverty and Famines,* economist Amartya Sen[11] offers an analysis of earlier famines in Bengal, Ethiopia, and the Sahel. A puzzling statistical pattern emerges, similar to the one just witnessed above for the most recent famine in Ethiopia. In fact, Sen cites substantial evidence to suggest that, in most famines occurring during the preceding century (for which reliable data are available), this anomalous statistical pattern can be found: despite natural disasters and "crop failures" that are alleged to disrupt the food supply, and irrespective of population size and growth or the condition of the environment, famines occur under what, at the time, seem to be favorable overall supply and production conditions. Famines occur, even though the local food supply seems adequate, in the view of local government and international experts at the time, to feed the population suffering from famine. Indeed, in many cases (as in Ireland during the Potato Famine, or in Ethiopia in 1974) food is actually exported from the "devastated" regions.[12]

On Sen's analysis, famine is more a function of specific and localized microeconomic factors than it is a function of broad, global parameters like weather and population patterns, or total food supply. Sen cites, by contrast, such problems as a sudden collapse in the normal economic abilities of specific vulnerable poverty groups in these regions—for example, nomadic herdsmen in the Sahel, or landless agricultural laborers and domestic servants dependent solely upon wage employment in Ethiopia—either to produce food for themselves (direct entitlements) or to exchange goods and services for food, or for wages to procure food (exchange entitlements). That is to say, famines occur among specific groups in a region whose direct or exchange-entitlement economic situation collapses. And this phenomenon cannot be adequately encompassed by, nor corre-

[10] E.g.: "Drought Spreads Death across Africa," *The New York Times* (November 4, 1984), p. 3; a Public Broadcasting Service "Frontline" documentary program, "Cry Ethiopia, Cry" (November 27, 1984); "Famine's Howl Echoes in Ethiopia," *The Los Angeles Times* (December 30, 1984).

[11] *Poverty and Famines: An Essay on Entitlement and Deprivation* (New York: Oxford, 1981).

[12] Cf. *Poverty and Famines,* chs. 6, 7.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

lated with, the behavior of the overall population, the weather, or a nation's overall food supply.

Although these microeconomic disasters may occur against the background of natural disasters, they are more directly a function of local political and economic policies: taxation, inflationary pressure (especially on food prices), unemployment patterns, governmental encouragement of developmentalist programs focused on "cash cropping," the absence of even minimal unemployment or welfare relief schemes, or border disputes in emerging nations (which inhibit traditional patterns of nomadic migration that themselves counter the effects of natural rainfall cycles).

For example, during the prolonged dry seasons and cycles of droughts that have characterized Sahel weather patterns for centuries at least, Sahel herdsman are accustomed to migrate in search of water and pasture for their animals. Following the colonial upheavals of the 1950s, however, these herdsmen now find that they are unable to migrate freely across relatively recently created national boundaries, and that their access to traditional "crisis" pastureland and water has been denied by fenced government farms where cotton is grown for foreign export. Moreover, they now find themselves obliged to sell off some of their surviving emaciated animals in order to pay a small annual tax either in the country of their "official residence" or as a border tax. Not only are their herds depleted by such policies, but the taxation program in effect creates a panic market in poor quality livestock. The price they receive per animal is now driven to dangerously low levels: the value of their remaining herds, and their ability to raise money, is severely jeopardized. This simultaneous dual collapse in direct and exchange entitlements is doubly serious: for, contrary to assumption, these nomads do not subsist just by eating their own animals; they also normally sell some of them in order to obtain cereals, bread, and grain.

Under these twin circumstances of loss of direct entitlement due to animal deaths, and collapse of market price due to panic and poor quality, mass economic ruin and starvation will certainly befall both these tribesmen, and those in the local economy whose own livelihood depends immediately upon trade with these tribesmen. All will suffer disastrous economic collapse, and in the absence of any tax relief or emergency economic assistance (which impoverished countries do not have and presumably cannot afford), large numbers of the affected groups will starve to death. Their deaths will be caused, not by normal or even extreme weather patterns or by the advance of the desert, but by the crushing burden of these otherwise normal

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

economic and political policies. This is apparently what transpired in the Sahel region during the period 1972–74.[13]

During the prolonged drought in Ethiopia in the early 1970s, the first to feel the pinch were not the farmers themselves, but those who worked for them. In a drought, the first thing that occurs economically is that the farmer is forced to dismiss any hired laborers upon whom he may normally depend. There are no crops to plant or harvest, and so there is no work to be done or income anticipated to pay the wages. (In Ethiopia, this occasionally resulted in selling children or turning out the older or female members of even the close family to fend for themselves.[14])

The absence of crops to plant and harvest during the drought in Wollo, combined with the local recession in economic trade, left farm workers, field hands, and other wage-earning service personnel with no food or property of their own to starve to death. This seems to have been the case in 1984 as well, despite the fact that the province, not to mention the nation as a whole, produced overall a sufficient quantity of food to feed its population, and actually allowed Wollo to export and sell some of its surplus to other regions. Since the agricultural laborers had lost their jobs and had no money to pay for food, they died of starvation, while those who still owned food were forced to sell it elsewhere or suffer the same fate.

In the meantime, both before and during the crises of the 1970s and 1980s, government ministers in Addis Ababa and international economic experts in Geneva and Washington were busily monitoring weather patterns, import-export statistics, population growth, and overall crop yields for the nation as a whole. These statistics foretold of no emergency; indeed, all seemed well. These forecasters never saw the disaster approaching, and were slow to understand or to acknowledge the crisis that resulted.[15]

Ethiopia's border dispute with Somalia may have exacerbated the famine in the eastern Ogaden province in 1974; likewise, the government's struggle with insurgents in Tigre and Eritrea has doubtless intensified the suffering of civilians in those provinces during the 1980s.[16] But in 1974 and 1984, it was Wollo, with adequate rainfall,

[13] Cf. Sen, ch. 8.
[14] Sen, p. 102.
[15] This explains, in turn, another common phenomena: that government ministers seem perpetually unprepared for and surprised by the famine crisis, and begin by denying its occurrence. Naturally so: their data and their faulty causal hypotheses tell them that no crisis could possibly be taking place.
[16] Jack Shepherd, "Drought of the Century," *The Atlantic Monthly* (April 1984), pp. 33–40.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

food to export, and no significant effect from either military conflict, that suffered the most severely from famine. While thousands starved in relief camps and remote outlying areas, food (in 1974, at least) was to be found on the shelves of local stores. These famines were not a function of drought, or of military conflict, or of overpopulation, or of environmental collapse, or of any serious inadequacy in the food supply. While primitive roads and lack of modern transport vehicles and systems did hamper the relief effort, transportation and distribution systems were evidently sufficient to put food on store shelves in Wollo, and to export Wollo's surplus production for sale in other regions of the country, including the capital city of Addis Ababa.[17] Neither does this phenomenon reveal hoarding, corruption, or the sinister influence of the black market. Rather, this is entirely consistent with the normal operation of a free economy in a time of crisis. It is economic "need," rather than true human need, that governs the flow of goods and services in such a market. As Sen himself remarks, "Starvation deaths can reflect legality with a vengeance."[18]

IV

The foregoing analysis of the most recent occasions of famine in Ethiopia supports Sen's microeconomic assessment of famine, and tends to discount the causal theories of both developmentalists and neo-Malthusians. Faulty theories of causation lead in turn to ill-advised relief efforts during a famine crisis, and so bear moral significance.

As a typical scenario, religious and civic groups and national and international charitable organizations mobilize to raise funds and obtain food. The food must then be shipped by boat or plane, at high cost, consuming precious time. The local transport systems are frequently inadequate to handle such a huge influx of foodstuffs. Trucks and planes must be donated and brought in to transport the food. Meanwhile, thousands starve while the food rots on loading docks or is sold on the black market. The very presence of this surplus donated food serves to depress market prices for locally produced food, further damaging, instead of developing, the fragile agricultural base whose fragility and low productivity are also alleged to be a contributing cause of the famine.[19] Tremendous sums of

[17] Cf. Adam Wilson, "How Starvation is Fueled in the Corridors of Power," *African Business* (April 1984), p. 10.
[18] Sen, p. 166. Cf. the conclusions drawn in his more recent article, "Property and Hunger," *Economics and Philosophy*, IV (April 1988): 57–68.
[19] Cf. James Bovard, "Free Food Bankrupts Local Farmers," *The Wall Street Journal* (July 2, 1984).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

money, quantities of surplus food, and human energy are expended —or rather, wasted—to little beneficial effect.

The harsh alternative of the neo-Malthusians is equally beside the point. They criticize the ineffective and inadvertent destructive tendencies of relief efforts, but propose merely to do nothing until the crisis abates through death, or else they focus on extremely expensive environmental rehabilitation projects and poorly conceived programs of population control. Yet family planning habits and population growth, after decades of intervention and experimentation, have proven as unpredictable and intractable as weather patterns. The most obvious demonstration of its deficiency, however, is obtained by testing neo-Malthusian logic in a larger global context. By the neo-Malthusians' own standards, the most severe environmental damage and excess of population over regional carrying capacity occurs in regions like Southern California, New York City, London, and Tokyo. While there is a great deal of poverty and homelessness in these regions, none has suffered from famine in recent memory.

Likewise, applying the developmentalist logic to a different context, some of the most severe droughts and disruption of local agricultural production in recent years have occurred in eastern Iowa, Illinois, South Carolina, and Georgia. These droughts have caused tremendous economic hardship for farmers and others in the affected regions, but no famines have occurred.

What these regions provide their residents, in contrast to the poverty-stricken regions of Africa and the Indian subcontinent, is certainly not immunity from variations in the weather, from insects, from population growth, or (alas) from environmental destruction. Rather, they provide an elaborate economic safety net, a network of economic institutions designed precisely to prevent utter destitution and death in times of economic dislocation and hardship which such natural events inevitably produce. Of course, it is precisely the lack and the apparent unaffordability of such social security systems which defines the latter nations as "poor" or "developing" to begin with. Nonetheless, it is vital to recognize these economic differences as fundamental in the cause, and in the relief, of famine.

All of this suggests the following prescriptions. First, in terms of monitoring the onset of emergencies, we need to focus less on huge macroeconomic factors, such as national food supply and population. We must also monitor local prices, especially food prices, watching for any sudden inflation of these relative to low-scale wages, or conversely, any sudden collapse of normal purchasing power among the least well-off in the country. We should monitor unemployment patterns among target groups, such as agricultural

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

labor, and commodity prices for locally produced food, such as cattle in the Sahel. In short, we must begin to monitor microeconomic, as well as macroeconomic, data.

In addition, the accumulation of such microeconomic data must be translated into appropriate political and economic safety valves. For example, during normal cyclical variations in the weather, such as dry spells in the Sahel, nomadic tribesmen might be permitted temporarily, by prior convention, to migrate freely across national boundaries. Further, their cooperating governments might initiate a temporary "tax relief" program to exempt these nomads from the usual annual taxation. It is clear that such modest and relatively affordable measures would have vastly reduced the extent of the Sahel disaster during 1974. The affected governments would, of course, require international monetary assistance to offset the loss of revenue, but this would be far easier logistically and less expensive financially to supply than the monumentally expensive and difficult relief efforts heretofore attempted. Likewise, the local governments might be assisted financially to establish a temporary "price support" system for the purchase of livestock, assuring the herdsman of a market for their cattle at a base price that would enable them to survive.

The possibilities are even more dramatic in Wollo. Those local landowning farmers with a modest reserve of grain to support their families, or the shopkeeper with food on the shelf, cannot themselves long afford to subsidize their less fortunate unemployed neighbors without risking death themselves under prevailing circumstances. Instead of the expensive and logistically infeasible provision of surplus food from other parts of the world, international relief agencies could intervene directly to achieve the necessary economic redistribution of locally available supplies. To accomplish this, the organizations might take the funds that they had collected for emergency relief and establish, in effect, a temporary welfare unemployment or "food stamp" program in the afflicted regions. In this region of the world, where the annual income is less than $100 per year, a little such assistance would go a long way. Such programs could be established in a matter of days at a fraction of the cost of conventional relief programs. In addition, these would entail far fewer administrative and logistical nightmares than currently accompany the practice of sending surplus food: for example, no black market; no food rotting on the docks; no collapse of local agricultural prices; no new roads to build; and very few trucks to be donated.

Most importantly, such programs, by restoring temporarily the collapsed exchange entitlements of the unemployed, would stimu-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

late, rather than disrupt, the local and regional economies and provide financial incentives that would restore and even develop the agricultural sector, while maintaining or even creating employment. The local shopkeeper would have paying customers again. Food already available locally would be able to be sold locally to those who need it most, rather than being exported to the cities. Farmers in other regions unaffected by the drought, in a nation whose overall food supply is more than adequate to its population, would once again have a regional market for their products and incentive to produce even more.

The myths regarding the causes of famine, and our customary responses to the tragedy of famine, are fostered by our own economic and political interests.[20] What these interests are, and why they are so intractable, would be the topic for another conversation. Nonetheless, if we are truly interested in saving the lives of those who urgently need our assistance, then it is necessary that our myths and our customary responses to this ancient scourge be set aside.

GEORGE R. LUCAS, JR.

Clemson University

[20] See my "Political and Economic Dimensions of Hunger," and also James Bovard, "Free Food Bankrupts Local Farmers," for a discussion of PL 480 and the benefits of the conventional international aid system to the donors, rather than primarily to the victims of famine. See also "Responses to Famine," an editorial in *The New York Times* (November 4, 1984).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

FAMINE AND DISTRIBUTION*

IN his interesting paper, "African Famine: New Economic and Ethical Perspectives," George R. Lucas, Jr. has reminded us that, in addition to "supply" and "demand," there is a thing called "distribution." Following Amartya Sen, he argues that the predominant cause of famine in sub-Sahara Africa is poverty, which in turn is a function of specific and local microeconomic factors. Change those factors, Lucas suggests, and the famine can be averted or remedied. Although I generally concur with his analysis and endorse the prescriptions made on the basis of that analysis, I think there are some flaws in his essay.

(1) Although Lucas suggests some new ways to accomplish the goal of alleviating famine, he does not offer any justification, grounding, or argument for the assumption that we ought to assist African famine victims. Nor does he justify the accompanying implicit assumption that the providing of such assistance is of a high priority. Contrary to the title, there appear to be no new ethical perspectives given in this paper.

(2) Who is Lucas criticizing? Who are the green-revolution developmentalists and the Hardinian neo-Malthusians today who are attributing famine in sub-Sahara Africa solely to undersupply or overdemand? Lucas's characterization of the developmentalist school seems to be at least a decade out of date. Post-green-revolution developmentalism, though still concerned with increasing the food supply per capita, is far more sensitive to the need for low-resource agriculture and for long-term sustainability of food-crop production practices. So, too, for the neo-Malthusian school; contemporary demographers directly concerned with reducing rapid population-growth rates actively advocate microeconomic intervention, especially for such things as the empowerment of women. Things have changed among both the "supply" and the "demand" schools since Sen's 1981 book.[1]

(3) Lucas (and to some extent Sen) is in danger of committing the same error he criticizes in others. He attempts to reduce all factors contributing to a food-crisis situation to one dominant cause, a failure of entitlement distribution. To the extent that all environmental,

* Abstract of a paper to be presented in an APA symposium on African Famine, December 29, commenting on a paper by George R. Lucas, Jr., this JOURNAL, this issue, 629–641.

[1] *Poverty and Famines: An Essay on Entitlement and Deprivation* (New York: Oxford, 1981).

0022-362X/90/8711/642–3 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

macroeconomic, infrastructural, production, political, and demographic questions are dismissed as irrelevant, or as contributory or exacerbatory, then a new "myth" has arisen, the distribution myth. Like the developmentalist's "food availability decline" and the neo-Malthusian's "exceeded carrying capacity," this new "myth" can distort extremely complex situations by oversimplification and the use of misleading metaphors and models.

(4) Finally, even granting that neither "undersupply" nor "overdemand" are necessary conditions for spot "famine" (since counter-examples exist), they are not thereby irrelevant to food crises. Long-term chronic food-quality shortages and nutrition-deficiency related deaths can exist in massive proportions among a population that is not technically undergoing a "famine." At best, Lucas has shown that undersupply and overdemand are not the predominant cause of "famine" in sub-Sahara Africa. He has not shown that they are not crucial causal factors (along with distribution) in the more normal hunger crises throughout the developing world. So, if they are "myths," they have only been shown to be so in this limited context; they have not been debunked once and for all, nor have the prescriptions for action that are based upon them.

WILLIAM AIKEN

Chatham College

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE VANISHING SUBJECT OF CONTEMPORARY DISCOURSE: A PRAGMATIC RESPONSE*

Charles Sanders Peirce's "Comment rendre nos idées claires" was published in the *Revue Philosophique* in 1879, a year after the English version appeared in the *Popular Science Monthly*. "How to Make Our Ideas Clear" was, of course, the article in which Peirce first put forth the doctrine of pragmatism. In 1905, Peirce wrote to the author of "La Notion de conscience," informing his good friend William James: "When you write English . . . I can seldom at all satisfy myself that I know what you are driving at. . . . But now that you are tied down to the rules of French rhetoric, you are perfectly perspicuous; and I wish . . . that you would consider yourself so tied down habitually."[1] James replied to this suggestion by confessing: "Your encouragement to me to become a French classic both gratifies and amuses. *I* will if *you* will—we shall both be clearer, no doubt. Try putting your [categories of] firsts, seconds, thirds in the Gallic tongue and see if you don't make more converts!" (*ibid.*, p. 435).

Whereas each of the cofounders of pragmatism believed it might be profitable for the other to translate his philosophical thoughts into the French language, today the reverse course is needed. For, in the interest of clarity and coherence, it seems necessary to translate some of the key figures in contemporary French thought into a more pragmatic idiom. Such, at least, is one way of stating the thesis of this paper. The trope of translation is misleading, however, to the extent it suggests an essential equivalence underlying diverse articulations; for translating the thought of such authors as Jacques Lacan, Michel Foucault, Jacques Derrida, and François Lyotard into the idiom of Peirce, James, John Dewey and G. H. Mead results in nothing less than transforming this thought. Hence, it is safer to speak of a pragmatic response to the dominant emphases of postmodern thinkers. The heart of this response is a critical appropriation of the fundamental insights of postmodern discourse, in particular, the insights into subjectivity. This appropriation itself is the achievement of a self-consciously historical and embodied agent whose status as such needs to be made evident, not impossible. Postmodern discourse tends to obscure and even to deny human agency as an enduring form of personal presence in the kaleidoscopic engagements of

* To be presented in an APA symposium on The Vanishing Subject of Contemporary Discourse, December 29. John J. Stuhr will comment; see this JOURNAL, this issue, 656–7.

0022-362X/90/8711/644–655 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

situated subjects. In contrast, the pragmatic perspective offers the conceptual resources to discern just this aspect of our subjectivity.

A moment ago I spoke of the 'pragmatic idiom'. What I mean by this expression is the philosophic language crafted by the classical American pragmatists (Peirce, James, Dewey, and Mead) *and* the philosophic sensibility expressed by this unique idiom. At the center of this sensibility is the felt sense that all discourse is a form of praxis and, in turn, all forms of praxis are ways finite, fallible agents take part in complex, uncertain circumstances. These forms of practice are to be seen as interactions between the human organism and its environment. This biological model is intended to render thought about ourselves and our engagements concrete and practical, not to warrant biological reductionism.

For the pragmatists, discourse transforms dumb or inarticulate creatures into thinking and knowing animals. By 'discourse', Peirce, Dewey, and Mead meant not only public exchanges but also "that preliminary discourse termed thinking."[2] Discourse here should not be taken too narrowly, however; in particular, it should not be conceived exclusively in terms of either the phonic or graphic symbols of natural languages. If 'language' is identified with such symbols, there is undoubtedly thought apart from language; however, "if 'language' is used to signify all kinds of signs and symbols, then assuredly there is no thought without language."[3] Accordingly, "thought lives, moves, and has its being in and through symbols, and, therefore, depends for meaning upon context as do symbols" (*ibid.*, p. 91). insistence upon the inseparability of thought and symbolization and, in addition, of symbolization and context is a defining feature of what I am calling the pragmatic idiom.

Context includes the enveloping situation and the selective interest of situated agents. Our cognitive struggles and the formal procedures and codified practices growing out of these struggles have their basis in what Dewey called problematic situations. For Dewey, it is a disclosure of our experience that "all reflection sets out from the problematic and confused. Its aim is to clarify and ascertain" (EN, 60). Whereas *all* reflective activity grows out of problematic situations, reflection can ascend to ever higher levels of generality and thereby become ever more attenuated from the situation out of which it arises. In other words, "[t]hinking takes place in a scale of degrees of distance from the urgencies of an immediate situation in

[2] Dewey, *Experience and Nature*, in *The Later Works of John Dewey, 1925–1953, vol. I* (Carbondale: Southern Illinois UP, 1981), p. 132. Hereafter cited as EN.

[3] John Dewey, *On Experience, Nature, and Freedom*, Richard J. Bernstein, ed. (Indianapolis: Bobbs-Merrill, 1960), p. 90. Hereafter ENF.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

which something has to be done" (ENF, 104). The more remote thinking is from such urgencies, the more likely it is to convert a provisional and legitimate disregard of context into a virtual and illicit denial of this factor. Of course, "philosophic thinking is, upon the whole, at the extreme end of the scale of distance from the active emergency of concrete situations. It is because of this fact that neglect of context is the besetting fallacy of philosophic thought" (ENF, 105). The pragmatic idiom is designed to counter the omnipresent tendency to commit this fallacy.

The constellation of French authors previously identified has exerted a considerable influence on Anglo-American thought, at least in certain circles within various academic disciplines, including philosophy. At or at least near the center of this influence is the decentering of the subject. In order to comprehend this, it is necessary to make explicit the conception of the subject bequeathed by Descartes to modernity; for the contemporary decentering of the subject is, in essence, the displacement of the Cartesian cogito and of the various disguised forms assumed by this ghostly figure in subsequent thought (e.g., in Locke, Kant, and Husserl).

In order to evaluate this decentering, it is helpful to draw attention to the efforts of the American pragmatists to make the communal practices of embodied agents, rather than the solitary reflections of a disembodied consciousness, the primary data of philosophical inquiry. This orientation of the pragmatists is perhaps best articulated by Dewey when he asserts:

> There exists at any period a body of beliefs and of institutions and practices allied to them. In these beliefs there are implicit broad interpretations of life and the world. These interpretations have consequences, often profoundly important. In their actual currency, however, the implications of origin, nature, and consequences are not examined and formulated. The beliefs and their associated practices express attitudes and responses which have operated under conditions of direct and often accidental stress. They constitute . . . the *immediate* primary material of philosophical reflections (ENF, 106).

The efforts by the classical American pragmatists anticipated the attempts of the contemporary French philosophers under discussion to offer an alternative paradigm to the dominant framework of modern thought, a framework in which the isolated individual is granted, in effect (if not also in intent), absolute authority as the source of meaning and the arbiter of truth.

Accordingly, two tasks confront us. We must discover what is generally meant by 'the decentering of the subject'. We must also, by means of a contrast to an analogous undertaking by the American

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

pragmatists, evaluate the way the cogito has been displaced by Foucault, Lacan, Lyotard, and Derrida. Given the scope of this paper, I shall focus primarily on just one representative figure from each camp, namely, Derrida and Dewey.

Before undertaking either of these tasks, however, two preliminary points need to be made. First, it is, of course, the case that the contemporary French writers with which I am concerned do represent a monolithic position. The differences among them are numerous and significant. Nonetheless, they are united in their opposition to the privileged status accorded the Cartesian cogito in post-Medieval thought and, in addition, in their debt to Ferdinand de Saussure's conception of language as a system of differences. Accordingly, it seems legitimate to consider an overlapping concern and, indeed, the collective influence of these diverse figures.

Second, since I shall present a pragmatic response to these French thinkers (i.e., a response rooted in what is, after all, the most distinctively "American" philosophic tradition), I would like to address the possible suspicion of cultural chauvinism. In my judgment, the influence of Foucault, Lacan, Lyotard, Derrida, etc., is not a malignant growth we need to remove from the body philosophic of this country, by means of surgical techniques developed by American *savoir-faire!* There is much of value in the writings of these thinkers. Indeed, these writings open up fields of inquiry or, at least, suggest strategies of investigation (e.g., Foucault's "archaeological" analyses of various discursive practices) it would be worthwhile for philosophers in this country, pragmatists and otherwise, to consider. Nonetheless, the most important insights of these French thinkers concerning the decentering of the subject might be more adequately formulated in a pragmatic idiom than in their own. Such a formulation enables us to render concrete and specific what far too often—and all too frustratingly—is left abstract and unspecified in the writings of the most influential contemporary French writers.

I. DECENTERING THE SUBJECT (SUBTEXT: OVERTHROWING THE FATHER)

As every professional philosopher knows, Descartes is traditionally recognized as the father of modern philosophy. As Richard J. Bernstein[4] has observed, however: "If we are to judge by philosophy during the past hundred years, this title can best be understood in a Freudian sense. It is a common characteristic of many contemporary philosophers that they have sought to overthrow and dethrone the father" (*ibid.*, p. 5). This is perhaps nowhere more apparent than in the various attempts to displace the Cartesian subject from its authoritative position.

[4] *Praxis and Action* (Philadelphia: Pennsylvania UP, 1971).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Descartes's triumph over skepticism depends on a particular conception of the human subject. The "I" who asserts "Cogito, ergo sum" is a disembodied, solitary, and theoretical consciousness. To be sure, at the end of the *Meditations on First Philosophy,* Descartes has regained his body, having established the existence of an Other and thereby having overcome his solitude in the "Third Meditation." At the beginning of the "Second Meditation," however, he proves his own existence as a thinking thing whose whole essence is to be a disengaged spectator and whose links to his own body, other bodies, and other persons are merely contingent. It is worthy to note that a disembodied consciousness is not likely to possess a gender identity; thus, the propriety of referring to Descartes as "he" is highly problematic.

Most thinkers today are highly skeptical about conceiving the subject in the manner of Descartes (about the principal move in his attempted refutation of skepticism). The contemporary French authors identified previously are prominent among these thinkers. Indeed, it is arguable that this constellation of authors is responsible for nothing less than a paradigm shift within a significant portion of our own academic community. Above all else, this shift has been *from* a paradigm centering on consciousness *to* one focusing on language; or, more fully stated, *from* a heuristic framework in which the most privileged position is accorded to consciousness conceived as a dynamic, intentional subject *to* an alternative framework in which an analogous position is granted to language defined as a system of differences.

In this new paradigm, conscious subjectivity (the "I" aware of itself as "I") is not treated as given at the outset of human existence; rather it is seen as constituted only in and through language. This is taken to imply that the origin of meaning is not to be traced to the acts of an intentional consciousness; rather, the origin of subjectivity is to be traced to the workings of symbolic systems. As Jonathan Culler[5] explains, the self is, in this view, "deprived of its status as source and master of meaning, [and as a consequence] it seems more and more like a construct: a result of systems of convention" (*ibid.,* p. 33). The "I" is not something given prior to discourse but is called into being as one who is addressed by others; and the stability of the subject (the fixity of the "I") is largely, if not totally, illusory, since the identity of the subject is essentially a series of positions assumed in response to the discourses of others. The impact of this way of understanding subjectivity has been considerable; it is evident in a

[5] *The Pursuit of Signs* (Ithaca: Cornell, 1981).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

variety of fields, including literary theory, literary criticism, social science, feminist theory, psychoanalysis, and philosophy.

Another important and influential aspect of this shift concerns calling into question the capacity of language to *refer* to a world not of its own making. The intention is not to deny the obvious but to reveal the concealed. We do obviously use language to refer to the world, but the referents of our utterances are determined by the language we use; in addition, what counts as a referent in one context can function as a sign in another. This point is characteristically made in terms of Saussure's analysis of the sign. In this analysis, the linguistic sign (like all other signs) is an arbitrary correlation of a signifier and a signified. But the status of anything as a signified is not fixed once and for all; in fact, it turns out to be essentially unstable, so that what is at one moment a signified can readily become at a later moment a signifier. For example, when a person makes the request 'Please get me a cigar', a cigar is the object being signified; but this object can itself function as a signifier of, say, power or taste. Sometimes a cigar may be just a cigar; but a cigar is always potentially more than a cigar. The potentiality of anything whatsoever to operate as a sign within some systems of signs needs to be not merely acknowledged in an abstract way but suspiciously considered in any concrete circumstance.

At what point, if any, is it possible to arrest the process by which signifiers are transformed into signifieds? The supposition that there *is* a point at which the play of signifiers stops has been a controlling assumption of Western thought. From the very nature of the case, however, such a terminus of signification is impossible; or, at least, so it is argued by Derrida and others.

To state the matter paradoxically, contemporary discourse is, in effect, *about* contemporary discourse in the sense that the function of texts is not to make present what is absent but to exhibit the processes and mechanisms of language itself. It is about turning the "about-ness" of language inside out! (Or is it outside in?) Language is not a mirror to be held up to either nature or humankind but a labyrinth from which we can never escape. It is, indeed, a labyrinth of labyrinths, so that the escape from one labyrinth is always an entrance into another. The longing for a transcendental signified is essentially the desire for an escape from the linguistic labyrinth.

At this point, let me call attention to a statement by Derrida concerning the status of the subject vis-à-vis language. My justification for focusing upon Derrida is simply that he is among the most influential of the thinkers under consideration and (in terms of the issue at hand) provides an excellent foil for Dewey (the pragmatist to be highlighted in the next section). In an interview with Julia Kristeva,

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Derrida[6] observes: "Language, and in general every semiotic code . . . are therefore effects, but their cause is not a subject, a substance, or a being somewhere present and outside the movement of differance" (*ibid.*, 28). He goes on to assert:

> Nothing—no present and in-different being—thus precedes *differance* and spacing. There is no subject who is agent, author, and master of *differance*, who eventually and empirically would be overtaken by *differance*. Subjectivity—like objectivity—is an effect of *differance*, an effect inscribed in a system of differance. . . . [I]ntuition, perception, consummation—in a word, the relationship to the present, the reference to a present reality, a being—are always *deferred*. . . . [T]he subject, and first of all the conscious and speaking subject, depends upon the system of differences and the movement of *differance*, that the subject is not present, nor above all present to itself before *differance*, that the subject is constituted only in being divided from itself, in becoming space, in temporizing, in deferal; and it confirms that, as Saussure said, 'language [which consists only of differences] is not a function of the speaking subject' (*ibid.*, pp. 28–9).

In light of the above considerations, there is a twofold sense in which to understand the title of this paper. The vanishing subject of contemporary discourse may refer to the effective displacement of the dynamic, intentional subject so central to the dominant humanistic tradition of modern European philosophy. It may also refer to the subject matter of contemporary discourse (for this discourse has become more and more a self-referential affair, a series of texts whose exclusive preoccupation is with other texts). On the one hand, the allegedly stable identity of the self has been dissolved into a series of positions into which the self is called and, indeed, by which the subject is constituted, positions assumed in response to the addresses of others: *who* I am has been and continues to be determined by how I am addressed by the Other, both other human beings and my own immanent other (i.e., the unconscious). On the other hand, the commonsensically assumed link between language and reality has been dissolved into a play of signifiers. In this play, the seemingly stable referents of everyday practical discourse transform themselves into signifiers whose status as signifiers is visible only to those who have adopted a hermeneutics of suspicion. This twofold "loss" has resulted in important gains, the most important being a deepened sense of the symbolic character of human existence and a radical suspicion regarding the dominant forms of discursive practice. Moreover, part and parcel of this "loss" has been a gain in insight into the historical contingency of all symbolic systems.

[6] *Positions*, Alan Bass, trans. (Chicago: University Press, 1981).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

But for these gains to be truly effective do we not need to abandon a perspective that appears to be systematically hostile to both the enduring or continuous character of human subjectivity and the denotative or referential function of human language? That is, for these insights to be practically significant do we not need to reject the very framework in which they are put forth? "Philosophers have exhibited proper ingenuity in pointing holes in the beliefs of common sense, but they have also displayed improper ingenuity in ignoring the empirical things every one has; the things that so denote themselves [in our lived, practical experience] that they *have* to be dealt with" (EN, p. 374). Do we not have a dramatic display of improper ingenuity by those authors who refuse to take seriously either the enduring character of the subject or the denotative function of language?

II. RECOVERING THE AGENT

As noted above, pragmatists no less than structuralists and poststructuralists have been committed to a radical critique of Cartesian subjectivity. Even so, the pragmatic decentering of the subject is markedly different from the structuralist and poststructuralist attempts to displace the cogito from its position of privilege and authority. Above all else, this difference concerns the status of the self as agent, a status rendered problematic due to the emphasis on the structures and processes by which subjectivity is constituted. This emphasis is found in the writings of the pragmatists no less than in those of the structuralists and poststructuralists; consequently, the question of agency arises equally for pragmatists, structuralists, and poststructuralists. Whereas Foucault, Lacan, and Derrida leave the status of the self-as-agent in doubt, however, the thought of Peirce, James, Dewey, and Mead drives toward the recognition of selves as sources of innovation and resistance (in a word, as agents). Since the case is made most easily in reference to Dewey, I shall limit my attention to his writings; however, I am confident that, on this issue, there is a consensus among the chief representatives of American pragmatism.

In 1940, this JOURNAL published an article by Dewey entitled "The Vanishing Subject of the Psychology of James."[7] This piece opens with the observation that:

> There is a double strain in the *Principles of Psychology* of William James. One strain is official acceptance of epistemological dualism. According to this view, the science of psychology centers about a *subject* which is 'mental' just as physics centers about an *object* which is material. But James's analysis of special topics tends, on the contrary, to reduction

[7] XXXVII, 22 (October 24, 1940): 589–599.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

of the subject to the vanishing point, save as 'subject' is identified with the organism, the latter, moreover, having no existence save with environing conditions (*ibid.*, p. 589).

Dewey proposes that we take the subject of experience to be human organism in continuous transactions with environing circumstances. For such subjects, there is no "problem" of how to establish the existence of a world beyond their own consciousness. Indeed, "if anything seems adequately grounded empirically it is the existence of a world which resists the characteristic functions of the subject of experience; which goes its way, in some respects, independently of these functions, and which frustrates our hopes and intentions" (ENF, 36).

It is as a consequence of the omnipresent capacity of the environing world to resist and even frustrate the exertions of the organism that the organism *is,* in one sense, a subject at all; for it is by virtue of this capacity that the organism is *subjected to* (as the etymology suggests, thrown under) factors and forces beyond its control. It is as a consequence of the limited but real capacity of the organism to counter resistances and to overcome frustrations that the organism is, in another sense, a subject. These two senses are explicitly distinguished in Dewey's *Experience and Nature.* Before turning directly to the way Dewey formulates this distinction, however, an indication of the context would be helpful.

Chapter 6 of *Experience and Nature* is entitled "Nature, Mind and the Subject." It opens with this observation: "Personality, selfhood, subjectivity are eventual functions that emerge with complexly organized interactions, organic and social" (EN, 162). Near the conclusion of this chapter, Dewey explores the importance for modernity of the distinction between subject and object, noting that it was this distinction, above all others, that provided the modern European mind with a way of interpreting its most important frustrations. The discussion of this importance provides the immediate context in which Dewey formulates the distinction between the two most legitimate senses of subjectivity. In this discussion, he wants to show how the dominant interpretive framework of the modern European mind is, in one respect, a local and provincial outlook and, in another, an important and indeed invaluable perspective. "Taken absolutely, the interpretation [of our experience, especially our experiences of frustration] on the basis of opposition of subject and object has no advantage over the other doctrines [formulated in the history of thought, e.g., the ancient Greek interpretation in terms of being and becoming and of matter and form]; it is a local and provincial interpretation" (EN, 184). Indeed, this interpretation is often marred by

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

an absurdity from which the other schemes are free; for subject and object have often been "antithetically defined" in such a way that they "can have logically no transactions with each other." Taken in reference to our efforts to overcome oppositions and to reduce the prospects of frustration, however, the interpretation of our experience in terms of subject and object (so long as these are conceived dialectically) is not only intelligible but also "more valuable than other modes of statement."

In this context, Dewey defines 'object' in a twofold manner and 'subject' in a similar way. An object is, in one sense, "that which objects, that to which frustration is due"; in another sense, it is "the objective; the final and eventual consummation, an integrated secure independent state of affairs." "The subject is that which suffers, is subjected [thrown under] and which endures resistance and frustration; it is also that which attempts subjection of hostile conditions; that which takes the immediate initiative in remaking the situation as it stands" (EN, 184).

The subject is, in other words, an agent-patient. This means that the subject who is susceptible to suffering, resistance, and frustration is not other than the subject who is capable of resistance, exertion, and innovation. These are distinguishable but inseparable aspects of those continuous, though continuously changing, centers of experience and activity. Although the emergence of these centers is only eventual, it is also eventful. Although the context of their emergence is a social setting inclusive of or (more accurately) constituted by discursive practices, what emerges from such settings is a unique, enduring center of action—in short, an agent.

To see discourse as a form of communal practice and, in turn, to practice no less than experience as an interaction between organism and environment is one way of decentering the subject without ingeniously ignoring "the empirical things that every one has; the things that so denote themselves that they/have/to be dealt with"—namely, the subject as agent-sufferer. "Respect for the things of experience alone brings with it such respect for others, the centers of experience, as is free from patronage, domination, and the will to impose" (EN, 392). Perhaps 'respect' is too weak a word here: 'reverence' would be more apt, for it suggests the most appropriate attitude toward what is, or should be treated as, inviolable. This attitude demands recognition of those centers of experience whose status as agent-sufferers is a direct, inescapable disclosure of our own lived experience, in particular, of our face-to-face encounters. Only it provides a counter to our tendencies to patronize and dominate others; indeed, only reverence secures a context in which a herme-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

neutic of suspicion can avoid degenerating into some form of nihilism.

III. CONCLUSION

Let me summarize the main points of my pragmatic response to the vanishing subject of contemporary discourse. The efforts of such postmodern thinkers as Foucault, Lacan, Lyotard, and Derrida are intelligible only in reference to what these thinkers take to be the dominant paradigm of Western modernity. In place of the paradigm centering on consciousness conceived as a dynamic, self-directed subject, these thinkers defend one focusing on language understood as a self-contained yet self-de(con)structive system of differences.

There is no question about the value of moving beyond subjectivism. There are, however, important questions about the alternative presented. This alternative is best judged only in comparison to the pragmatic attempt to displace the Cartesian subject from its authoritative position in Western thought. In the writings of Peirce, James, Dewey, and Mead, we encounter a critique of the cogito no less radical than that found in the writings of Foucault, Lacan, Lyotard, and Derrida. The principal thrust of the pragmatic critique is, however, more obviously and directly a recovery of embodied agents entangled in historical circumstances. Whereas the poststructuralists emphasize impersonal structures and processes, the pragmatists stress intersubjective transactions and practices.

This is especially evident in reference to language. For the poststructuralists, language is above all else a system of differences in which the play of signifiers, leaving traces of traces, is what ultimately counts; whereas for the pragmatists language is a set of practices by which embodied agents orient themselves to one another and establish shared frameworks of ongoing activity. Because of their view of language, the poststructuralists leave problematic the status of the self as agent; whereas the pragmatist—also because of their view of language—decenter the disembodied cogito precisely in order to recover the embodied agent. Both groups of thinkers, though in different ways, illuminate the practices, discourses, and institutions in and through which embodied agents make practical decisions. It is even arguable that the French postmodernists offer more detailed and subtle analyses and critiques of these practices, discourses, and institutions than do the American pragmatists. In these analyses and critiques, however, the self seems to be reduced to the interstices in a system of differences, a passive construct of social forces.

In contrast, the American pragmatists take their stand with common sense and refuse to relinquish the self-as-agent, the subject as source of innovation and resistance. Their uncompromising recogni-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tion of the historically and institutionally embedded character of human subjectivity does not prompt them to reduce the subject to a screen upon which conflicts and contradictions are projected; rather, their own lived experience prompts them to insist upon seeing human selves as embodied and, thus, embedded *agents*. In this clear and commonsensical insistence, the American pragmatists differ markedly from the French postmodernists; and here is a difference that truly makes a difference. At any rate, this is what I hope to have made plausible in this paper.

The pragmatic response to the vanishing subject of contemporary discourse should be one of critical appreciation, one seeing the limitations and even dangers of discursive practices systematically hostile to the denotative method, on the one hand, and the resources and power of such practices, on the other. What Dewey wrote about the course of philosophy in its entirety may, with equal justice, be said of this recent chapter in Western philosophy, namely,

> As I reflect upon the historical course of philosophy I am unable to find its course marked by notable successes in the matter of conclusions attained. I yield to none, however, in admiring [yet critical] appreciation of the liberating work it has accomplished in opening new perspectives of vision through its sensitivity to problems it has laid hold of in ways which, over and over again, have loosened the hold upon us exerted by dispositions that owe their strength to conformities which become so habitual as not to be questioned, and which in all probability would still be unquestioned were it not for the debt we owe to philosophers (ENF, 14).

The forms of domination embedded in and perpetuated by the dominant discourses, practices, and institutions of Western culture are more sharply and finely discerned by those whose "sensitivity to problems" has been heightened and deepened by exposure to the writings of Foucault, Lacan, Derrida, Lyotard, etc. The emancipatory power of the critical interventions undertaken in such works as *Discipline and Punish* and *The Post-Modern Condition* cannot be fully realized, however, apart from a pragmatic recovery of the embodied agent. This recovery shifts attention from the deconstruction of texts to the reconstruction of institutions; but it does so by critically appropriating the ways the deconstruction of texts might serve the reconstruction of institutions. For such service to be rendered, however, a "translation" of contemporary discourse into a more pragmatic idiom is required.

VINCENT M. COLAPIETRO

Fordham University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

SUBJECTS CONSTRUCTED, DECONSTRUCTED, AND RECONSTRUCTED*

What are the significant relations, if any, between French postmodernists and American pragmatists? What is the same or, at least, largely shared? What is different or, perhaps, even incompatible? What, if anything, do these two groups of philosophers have to offer to one another? And what does an explication of their relations, blood lines, and official papers have to offer to us?

These questions have not been raised frequently, examined urgently, or answered in depth. Philosophers operating within either one of these traditions largely have ignored the other tradition, even when they have been concerned with "the Generalized Other" or, more generally still, simply "the Other."

These questions are not new, however. Not surprisingly, they have been answered in a great many different ways for a great many different purposes. Some writers argue that the two traditions are marked by irreducible major differences. By contrast, others find overriding deep similarities. In "The Vanishing Subject of Contemporary Discourse: A Pragmatic Response," Vincent Colapietro takes a middle position, arguing that both contemporary French theory and classical American pragmatism decenter, displace, and replace Cartesian accounts of consciousness, subjectivity, and reference. The similarities stop here, however. The postmodernists deconstruct the Cartesian self by means of an account of language that renders problematic the self's agency and language's denotative function. Pragmatists, on the other hand, provide an equally radical but more successful critique of modern thought and culture. Their account of self, nature, and communication permits appropriation of the postmodernist insights for the purposes of cultural reconstruction, instead of mere textual deconstruction.

As an overview of the two traditions, this sort of "middle" view is largely correct and uncontroversial: both traditions reject Cartesianism, but in different ways for different reasons and with different theoretical and practical results. Of course, this view is not a "translation" of postmodernism into pragmatism. And if it asks pragmatists to grasp postmodernist analyses, it is primarily a call to post-

* Abstract of a paper to be presented in an APA symposium on The Vanishing Subject of Contemporary Discourse, December 29, commenting on a paper by Vincent M. Colapietro, this JOURNAL, this issue, 644–655.

0022-362X/90/8711/656–7
© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

modernnists to become pragmatists. This invitation to conversion has deep roots: John Dewey called pragmatism a Copernican Revolution and William James compared it to the Protestant Reformation, naming it the philosophy of the future.

This important project, however, is both insufficiently and overly cautious. It is insufficiently cautious in two respects. First, even as overview, it blurs basic differences among French theorists. This is a complex issue, as Michel Foucault noted, admitting that he did not see anything common to theorists identified as postmodern. For example, Jacques Derrida's post cards and play of signifiers differ crucially from Foucault's ethics and genealogies of subjects. Deconstruction of the subject versus its construction through historical practices: this difference should make a difference, especially to pragmatists, and it must be spelled out. Second, the messy relations between French theorists and American pragmatists cannot be captured by an analytical checklist of themes shared and unshared. For a project of appropriation, this is too general; it must become more historical and pragmatic—that is, more attuned to cultural context. A thoroughly pragmatic response to postmodernism must locate both traditions in terms of the (different) problems, times, and places in which they arose.

In this light, Colapietro's pragmatic response to contemporary French theory is also too cautious in two respects. First, pragmatists do not need to appropriate French decenterings of Cartesian subjects. Pragmatists did this first and did it better. Instead, they must explore postmodern accounts of subjects and cultural formations given traditional pragmatic commitments to humanism, liberalism, and community. Can postmodern theory join pragmatist politics? If so, how? This question demands a response to contemporary practices—media, prisons, schools, hospitals, families, churches—as well as contemporary discourse. This response, if genuinely pragmatic, cannot proceed by partitioning culture into two spheres and giving one—the public—to pragmatism and one—the private—to postmodernism. This strategy, now championed by Richard Rorty, sacrifices pragmatism itself, and so necessitates a detailed alternative.

JOHN J. STUHR
University of Oregon

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

CARTESIAN SKEPTICISM AND INFERENCE TO THE BEST EXPLANATION*

The problem of skepticism about the external world, or Cartesian skepticism, has its roots in the underdetermination of theory by evidence. We each adopt a body of common-sense beliefs about the world which answers to our sensory experience. In principle, however, the beliefs we base on that experience are subject to underdetermination, and we can devise radical alternatives to the common-sense account. Such alternatives take the form of skeptical hypotheses, like Descartes's fiction that his experiences are caused by an evil demon.

Certainly, when the choice arises, we hold to the common-sense view, and reject its skeptical competitors.[1] But what (epistemic) reasons can we have for doing so? In cases of underdetermination generally, principles of inference to the best explanation can license the choice of one theory over others. Accordingly, we would be justified in preferring the common-sense account to skeptical hypotheses, if the common-sense account provides better explanations of why our experience is the way it is.[2] My purpose here is to inquire into the explanatory advantages of the common-sense view, and to develop a response to skepticism along the lines just indicated.[3]

One obstacle to carrying out this project is that the standards by

* To be presented in an APA symposium on Cartesian Skepticism, December 30. Berent Enc will comment; see this JOURNAL, this issue, 667–8. I would like to thank Phillip Bricker, Anthony Brueckner, Jeremy Butterfield, Stewart Cohen, Richard Feldman, Mark Johnston, Peter Lipton, Thomas Tymoczko, and Daniel Velleman. Research for this paper was supported by the American Council of Learned Societies and the National Endowment for the Humanities.

[1] Some philosophers, especially followers of Ludwig Wittgenstein, would deny that skeptical hypotheses can genuinely compete for acceptance with the body of our common-sense beliefs. See, for example, Stanley Cavell, *The Claim of Reason* (New York: Oxford, 1979), pp. 218–220.

[2] This approach to skepticism has been advocated by Michael Slote, Frank Jackson, Jonathan Bennett, James Cornman, J. L. Mackie, and Alan Goldman, among others.

[3] On certain views about skepticism and about inference to the best explanation, this approach to skepticism will seem ill-conceived. One might hold that it is simply constitutive of rationality to reject skeptical hypotheses out of hand; thus, it is unnecessary to enter into the relative explanatory merits of the common-sense view and its skeptical alternatives. From another point of view, the explanatory advantages of the common-sense view could never give us a reason to accept it as *true*, rather than as merely handy or to our taste. The issues that arise here are important, and they must be addressed at some point by anyone who bases an answer to skepticism on explanatory considerations. These very general objections will be moot, however, if the appeal to explanatory considerations does not even succeed on its own terms. Whether it does so is my present concern.

0022-362X/90/8711/658–666 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

which explanations are evaluated are themselves difficult to identify and to make precise. In what follows, I shall be making some controversial assumptions about explanatory goodness, and I shall have to rely on largely unanalyzed notions of simplicity, ad-hoc-ness, and the like. To be explicit, I shall presuppose:

(a) Ad hoc explanations should be avoided, i.e., very roughly, if *A* is offered as an explanation of *B*, *A* ought not to be isolated from other explanations and data (it ought to be independently testable, it must figure in the explanation of something other than *B*, etc.).

(b) Other things being equal, a simpler explanation is superior to a more complicated one.

(c) Where explanation is concerned, more is better, if you get something for it. In particular, it is desirable to be able to give higher-level explanations of lower-level ones.

Another methodological point requires some comment. In comparing skeptical hypotheses with our everyday account of the world, I shall exclude from the latter any advanced scientific beliefs. To be sure, science adds great power and coherence to our explanations of phenomena, and one might argue that no explanatory scheme the skeptic devises could seriously compete with our best scientific theories. But it seems implausible that, without such theories, we would lack adequate grounds for rejecting skeptical hypotheses. Accordingly, I shall try to show that even a scientifically unsophisticated common-sense view of the world provides more adequate explanations than its skeptical competitors.

I

Our beliefs about the external world serve an explanatory function. A person's sensory experience exhibits patterns and regularities at many levels, and our common-sense beliefs account for these in ways that seem to be coherent and economical. I shall call the body of these beliefs the *real-world hypothesis* (RWH).[4]

The skeptic points out that there are alternative explanations of how a person's sensory experience arises. In principle, a great many ways of formulating and developing these counterhypotheses are open to the skeptic—for example, through various stories about evil

[4] This way of putting things may seem unfortunate to those who reject the representative theory of perception. But the point could be recast as follows: we have a set of beliefs about the world, i.e., the RWH. Our having those beliefs admits of alternative explanations, including skeptical explanations and the RWH itself. The tenability of skepticism turns on whether the truth of the RWH provides a better explanation than do skeptical hypotheses of why we believe the RWH in the first place.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

demons and brains in vats. But elaborate (not to mention crazy) fantasies of deception may be only tenuously connected to the content of one's experiences and may lack cohesiveness. For instance, suppose that you seem to see some snow falling, and the skeptic suggests that this experience is being foisted on you by a demon. Then, to explain why the demon makes you have snow experience (rather than experience of some other kind), the skeptic tells you that there is a second demon that has put the first one up to it. Clearly, we are not getting anywhere; positing a second demon that directs the first on this occasion (and does only that) is explanatorily idle or ad hoc. The skeptic could try to escape such a result by refusing to say in any detail how your experiences come about. A hypothesis in this vein might specify only that your experiences are all caused by some deceptive spirit, and no more. The cure is as bad as the disease, however: the skeptic will succeed in avoiding ad hoc higher-level posits only by foregoing higher-level explanations altogether.

The RWH, by contrast, gives us a rich and well-integrated explanatory apparatus. We not only posit objects that cause our experiences, we are also able to explain why and how these objects behave as they do. If the explanations provided by a skeptical counterhypothesis are either ad hoc or impoverished in comparison with those of the RWH, then we have good grounds for preferring the latter to the former. According to the skeptic, we fail to know things because the RWH is faced with competitors that we have no reason to reject. But we have just seen that not any competitor will do. The skeptic's position will be empty unless he can provide us with reason to think that a *satisfactory* competitor exists (in particular, a sufficiently rich competitor that is not unduly burdened with ad hoc explanatory posits).

The lesson here is that the skeptic needs to frame an alternative that matches the RWH very closely. If a skeptical hypothesis can be made sufficiently similar in relevant respects to the RWH, then, one might expect, that skeptical hypothesis will match the RWH in explanatory adequacy. To the extent that explanatory virtues like coherence, depth, and simplicity are matters of theoretical structure, a skeptical hypothesis that is isomorphic to the RWH will explain things just as well as the RWH does. An improved skeptical hypothesis of this sort has to satisfy two principal constraints: (i) it should invoke items corresponding to the elements of the RWH; (ii) it should also posit, as holding of these items, a pattern of properties, relations, and explanatory generalizations mirroring those of the RWH.

As an example of how this would work, suppose you seem to see the wind blowing a piece of paper off your desk. According to the

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

RWH, your visual impressions of the paper flying off your desk are caused by the paper. Similarly, your tactile sensations of the wind are caused by a real movement of air against your skin. And, finally, the wind stands in a relation of cause and effect to the movement of the paper. The skeptic's procedure will be to extract the explanatory skeleton or core from the RWH—that there are *some* entities bearing *some* properties that are related in ways exactly analogous to those specified by the RWH—and then to add that the entities and their properties are somehow different from the ones mentioned in the RWH.[5]

Thus, a skeptical hypothesis might present the following alternative explanation of your experiences. All that there is to the world is your brain in a vat, and a computer that is connected to your brain. Your tactile experiences are caused by the realization of a computer program that simulates wind, and your visual impressions are caused by the realization of another program that simulates a paper blowing off a desk. Also, the skeptical hypothesis can specify that the first routine calls the second, so that (as in the RWH) the cause of the wind experience would be the cause of the cause of the paper-blowing experience. This way of reconstructing the explanatory structure of a small fragment of the RWH might be extended to apply to all the entities and explanatory connections posited by the RWH. The result would be a skeptical hypothesis that was completely isomorphic to the RWH, with portions of the computer disk supposed to occupy the explanatory roles we normally assign to familiar objects.[6] I shall be calling this the *computer skeptical hypothesis* (CSH).

Of course, the CSH is an outlandish suggestion, and we are confident that it is false. Yet, in reflecting on this situation philosophically, it is possible to misread what has gone wrong. One proposal is that skeptical hypotheses are invariably burdened with more unexplained

[5] Basically, this amounts to something like forming the Ramsey sentence of the RWH and adding to it further specifications that, in each case, the object or property denoted by the bound variables is something other than the one posited by the RWH. The RWH itself can be construed as the "Ramsey sentence" plus the stipulation that the objects and properties called for by the "Ramsey sentence" are indeed the familiar ones. See here Grover Maxwell, "Theories, Perception, and Structural Realism," in R. Colodny, ed., *The Nature and Function of Scientific Thought* (Pittsburgh: University Press, 1970) and for some needed refinements, David Lewis, "How to Define Theoretical Terms," *Philosophical Papers,* vol. I. (New York: Oxford, 1983). The possibility of framing skeptical hypotheses with the same structure as the RWH is noted by Lawrence Sklar in his "Saving the Noumena," *Philosophy and Spacetime Physics* (Berkeley: California UP, 1985), pp. 59–60.

[6] We need not suppose that the computer itself was built or programed by anyone. Rather, this hypothesis is to be understood simply as a description of an alternative way the (physical) world might be.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

explainers than is the RWH.[7] The CSH will lack answers to questions like 'Why does the computer operate the way it does?' or 'Where did the computer come from in the first place?' But it is not at all clear that the RWH does any better in the face of analogous demands. Both the CSH and the RWH invoke ultimate regularities that are not themselves explained, and neither can account for the existence of the physical world as such. Generally, since the RWH and the CSH are meant to have the same structure, anywhere the RWH can explain a lower-level phenomenon by a higher-level regularity, the CSH should be able to do the same. The CSH will have unexplained explainers only insofar as the RWH has them also.

Another suggestion that enjoys some currency is that the RWH is, in a very straightforward way, simpler than the CSH, and hence to be preferred. The idea here is that there would be a one-one mapping from the objects posited by the RWH to their stand-ins in the computer's memory, where these are treated as discrete individuals. There are, though, items required (at least tacitly) by the CSH which escape this mapping, e.g., the computer's central processing unit and perhaps the brain in the vat itself. So, the argument runs, the CSH is committed to the existence of more items than the RWH, and is to be preferred on that account.

This line of thought is problematic in several respects. First, one could just as well argue that the CSH is simpler than the RWH, on the grounds that the CSH posits only two objects (the computer and one's brain), whereas the RWH is committed to the existence of a great many more things. Moreover, it is far from clear that, all by itself, positing fewer entities is a theoretical virtue.[8] And finally, if need be, the CSH could be revised to eliminate the role of the central processing unit altogether. The skeptic could suppose that the elements of the computer memory act directly on each other, and on the seat of consciousness, in causal patterns that mirror those of the RWH.

Now, as will emerge shortly, I think there is something right about the claims that the CSH is less coherent and less simple than the RWH. But if explanatory coherence and simplicity are treated solely in structural terms, it should not be surprising that these claims do

[7] A claim of this sort is made by Alan Goldman, although it is directed at a fantastical skeptical story that postulates experimenters with deceptive motives. See Goldman, *Empirical Knowledge* (Berkeley: California UP, 1989), p. 212.

[8] It could be objected that what matters for explanatory adequacy is not economy with respect to the number of individuals posited, but rather in the number of different kinds invoked. But this does not appear to help—the skeptic can get by with just a few kinds of things (brain, vat, computer) while the RWH might be said to invoke these and many more.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

not go through. After all, the causal-explanatory structures invoked by the RWH and the CSH are identical; the two differ only as to what entities bear the specified causal relations to one another.

The rejoinders just considered miss something important about the motivations behind the skeptic's argument. At root, the skeptic questions our ability to read off the "real" or intrinsic character of things from those things' causal behavior. This challenge emerges in its simplest form with the initial thought that one's experience of any familiar object might be caused by something other than that object (e.g., an evil demon). The point is that the known effect—namely, your experience—does not fix the character of its cause.[9]

On the face of it, the requirement that a skeptical hypothesis must have a more fully articulated structure—one that matches the RWH in various ways—seems insufficient to meet this problem. For, if we assume that causal relations are contingent and that there is in principle no obstacle to our positing whatever causal relations we like, what reason could there be why one set of entities is better suited than another to occupy the positions within the structure of the RWH itself? It would appear that, in principle, there should be skeptical hypotheses that will explain the contents of one's experience just as well as the RWH. The choice between such hypotheses and the RWH will then be arbitrary, giving the skeptic what he needs.

II

To appreciate the superiority of the RWH over its skeptical competitors, we need to take into account the content, as well as the form, of the explanations the RWH provides. In particular, our ordinary view of things involves beliefs in the existence of objects with familiar spatial characteristics (e.g., we believe that there are bricks that are oblong and oranges that are round). The ascription of specific spatial properties to objects does explanatory work within the RWH (e.g., accounting for why oranges roll easily and bricks do not). Since the CSH posits objects with altogether different spatial characteristics —we are assuming that its objects are just portions of a computer disk—the CSH will have to account for the relevant phenomena in some other fashion. But by bringing in these additional explanations (whatever they may be), the CSH runs the risk of taking on a more elaborate explanatory apparatus than the RWH. To put the point I am trying to make more directly: niceties aside, the fact that some-

[9] This was the way Kant understood the situation. The skeptic, he says, "assumed that the only immediate experience is inner experience and that from it we can only *infer* outer things—and this, moreover, only in an untrustworthy manner, as in all cases where we are inferring from given effects to determinate causes"; *The Critique of Pure Reason,* N. K. Smith, trans. (New York: St. Martin's, 1965), p. 245.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

thing is spherical explains why it behaves like a sphere (in its interactions with us and with other things). If something that *is not* a sphere behaves like one, this will call for a more extended explanation.

This intuitive claim is bound to raise some philosophical qualms. Why must the fact that the CSH invokes *different* configurations of matter in its explanations mean that CSH has to be *more complicated* than the RWH? Again, setting niceties aside, why is the skeptic not free to stipulate that, in his account, it is certain magnetic patterns on a disk, not spheres, which behave like spheres (at least in terms of the experiences they bring about, directly and indirectly)?

Let us see just what would be involved in maintaining an explanatory parity between the CSH and RWH. To fix ideas, suppose that, according to the RWH, there is a hyacinth beside your doorway. For each RWH object, there has to be a CSH counterpart, which we can imagine to be the piece of the computer disk which stores the information about the object to be simulated. So, the CSH would have it that there is a piece of the disk holding a file about a hyacinth beside your door, specifically. Moreover, wherever the RWH assigns a certain property to the hyacinth, the CSH must ascribe a corresponding, but different property to the hyacinth's CSH analog. According to the RWH, the hyacinth has a particular location, namely, that of being beside your door. The hyacinth counterpart will have some parallel feature, which we might call a "pseudo location." The pseudo location of the hyacinth counterpart is just that physical property in virtue of which the counterpart simulates being located near your door. In general, what the RWH explains by reference to genuine locations, the CSH will explain in terms of these pseudo locations.

Since we make reference to the locations of objects in giving various everyday explanations, location properties are part of the explanatory apparatus of the RWH. Now, we find that the (genuine) locations ascribed to any two objects at a time are invariably different. We do not need any empirical law or regularity to explain this; it is a necessary truth pertaining to the nature of physical objects that there cannot be two such objects at the same place at the same time.[10]

The explanatory structure of the CSH is meant to duplicate that of the RWH. Since the CSH is isomorphic to the RWH, and the RWH always ascribes different locations to the objects it posits, the CSH will invariably ascribe different pseudo locations to things it posits.

[10] For a discussion of this principle, see Denis Robinson, "Re-identifying Matter," *The Philosophical Review*, XCI (1982): 317–341; on the role of necessary truths in explanations, see Clark Glymour, "Explanation and Realism," in J. Leplin, ed., *Scientific Realism* (Berkeley: California UP, 1984), esp. pp. 184–6.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

This calls for an explanation, if possible. At this point, however, the CSH faces a loss in either simplicity or explanatory power. To make the issue more concrete, imagine that the way things work in the CSH computer is that each object's pseudo location is the physical realization of having coordinates (x,y,z) written in its file.[11] There will have to be some explicit principle within the CSH that no two objects are to be assigned the same pseudo location, i.e., that no two objects are to have the same coordinates written in their files. Otherwise, the fact that no two objects have the same pseudo location remains unexplained. Of course, the CSH would include within it the necessary truth that two physical objects cannot occupy the same *genuine* location in space, but this is of no help to the CSH in explaining why two of its objects cannot have the same *pseudo* location. To achieve this, it would appear that the CSH has to add an extra empirical regularity, to which no regularity in the RWH corresponds. Such an addition will make the CSH inferior to the RWH on simplicity grounds, however.

The skeptic could escape this outcome if it could be guaranteed by some other necessary truth that different CSH objects will have different pseudo locations. In other words, the pseudo location of a CSH object would have to be encoded by some physical property P (other than that of having some specified location), such that it is impossible for two physical objects to have P at the same time. But it seems to me that there are no such physical properties. After all, if a given physical object with whatever properties exists at one place, it appears perfectly possible for there to be an absolutely similar object elsewhere, instantiating all the same properties at the same time—except location.[12]

Actually, the problem facing the skeptic is a general one, independent of the fact that CSH itself invokes physical objects (i.e., bits of computer disk) in its explanations. Suppose that the skeptic offers instead a quasi-Leibnizian hypothesis, according to which the world consists solely of minds and their properties. These minds and their states are supposed to act in ways that mirror the behavior of everyday things as specified by the RWH. Each mind that stands in for a RWH object must have a property corresponding to the genuine location the RWH ascribes to its object; this pseudo location will be a (partial) mental state. The question arises again as to why these

[11] For purposes of exposition, I am pretending that an object is located at a point rather than a region.

[12] Of course, there are characterizations like 'the only building taller than 110 stories' or 'identical to Socrates', which are satisfied by at most one object at a time. If these expressions involve reference to properties, they are properties of a different type than those with which I am concerned here.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

pseudo locations are invariably different from one mind to another. Presumably (*pace* Leibniz), it is possible for two different minds to think exactly the same thing at the same time, so no necessary truth prevents them from having the same pseudo location. Once again, such an occurrence would have to be ruled out by some kind of extra "exclusion principle," for which no counterpart exists in the RWH.

I claimed earlier that our normal ascription of spatial properties to things does real explanatory work; furthermore, it seems plausible that you incur an added explanatory burden if you suppose that something lacking a particular spatial property still behaves as though it had it. What I have been saying about locations and pseudo locations makes this same point on a more abstract level. In skeptical hypotheses, some other property (e.g., a magnetic property or a mental property) is supposed to substitute for the spatial property of being located at particular place. As we have seen, further explanation is then needed to establish why these properties, which are not genuine location properties, behave as though they were. It seems that this sort of difficulty will attach to skeptical hypotheses generally, giving us good reasons to reject them.[13]

III

I have presented some antiskeptical arguments based on explanatory considerations. But surely there is a world of familiar objects about us, and we have known that all along. So what, then, is the point of giving these arguments in the first place? This question deserves an extended response, but for now a very brief answer will have to do. I take it that the specious character of the explanations the skeptic offers is immediately apparent—they come across as contrived or unduly indirect—and this is a reason why we reject skepticism as a doctrine. Realizing that skeptical hypotheses are defective, however, is not the same thing as spelling out precisely what their defects are. To do this requires philosophical work—work of the sort I have undertaken here.

JONATHAN VOGEL

Amherst College

[13] One might try to frame a skeptical hypothesis that avoids this difficulty by assigning to objects different locations (and spatial properties generally) in place of those specified by the RWH. Formulated this way, our problem becomes one of choosing a particular geometry of the world from among those logically compatible with the empirical data, and one might continue to defend the choice of the RWH by appeal to explanatory considerations. See here Lawrence Sklar, *Space, Time, and Spacetime* (Berkeley: California UP, 1977), pp. 91–101, although Sklar himself is highly critical of such uses of inference to the best explanation. Sklar has a valuable discussion of the affinities between Cartesian skepticism and problems in the epistemology of geometry.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

IS REALISM REALLY THE BEST HYPOTHESIS?*

Jonathan Vogel argues for the conclusion that the real-world hypothesis (RWH) is a better explanation of our experience than the skeptical hypothesis (SH). His argument moves in two stages; in the first stage, he imposes the requirement that any alternative hypothesis has to be structurally (syntactically) isomorphic to the RWH; in the second stage, he argues for the explanatory superiority of the RWH over any hypothesis which shares the structure of the RWH but which rejects its ontology. Vogel's reasoning here is that the RWH will avail itself of the metaphysical necessary truths that govern physical objects, whereas its competitors will have to write these in as additional empirical laws.

I propose to show that, in both stages, Vogel has been unfair to the skeptic.

(1) History of science provides us with examples of new theories which dethroned old ones and which were not structurally isomorphic to the old ones. (One example is Galilean versus Aristotelian mechanics.) In these cases, it would have been unfair to dismiss the new theory merely on the grounds that it failed the isomorphism requirement. Lack of resemblance to the structure of an established theory is not by itself sufficient to convict a theory, because each theory has a different way of marking off the boundaries of its proper explananda, of classifying them, and of evaluating the significance of "observables." In saying this, I am not recommending a Kuhnian view of science; I am merely noting that the logical-empiricist supposition about the existence of a pool of rock-bottom neutral observation sentences against which *all* theories are to be tested is a convenient oversimplification. And if two theories do not share the same observation sentences, they are unlikely to possess identical structures.

The above result about scientific theories has immediate application to epistemology: to assume that the experience to be explained by the SH is properly to be described in the language of the RWH is to be unfair to the SH. The inferential and learned nature of our perceptual mechanisms may already be programmed with a rudimentary form of the RWH; and the skeptic could consistently maintain that, once we "unlearn" that program, our experience will become very different. (Here the skeptic has a lot to learn from Berkeley,

* Abstract of a paper to be read in an APA symposium on Cartesian Skepticism, December 30, commenting on a paper by Jonathan Vogel, this JOURNAL, this issue, 658–666.

0022-362X/90/8711/667–8 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

who argued that the only rock bottom "givens" in our visual experience are colors and light intensities—distances, sizes, etc., being imposed on the given by "habit.")

(2) We can see how Vogel is being unfair to the skeptic in the second stage if we suppose a SH that *is* structurally isomorphic to the RWH. The argument here is reminiscent of contemporary debates between scientific realists and antirealists. It is true that when, in one interpretation of a scientific theory, physical objects are the referents of the theoretical terms, and on another interpretation, "fictitious" entities or "pseudo objects" become the referents, one can show, along the lines that Wilfrid Sellars, Hilary Putnam, and Richard Boyd have used and Vogel adopts, that the first interpretation is explanatorily superior. But I tend to think that more sophisticated versions of antirealism (e.g., Bas van Fraassen's) are immune to this criticism. Adopting one such version, the skeptic could reasonably maintain that his hypothesis "saves all the phenomena" that the RWH does, and it does this *without* introducing "pseudo objects," *without* giving any semantic interpretation to any of the nonlogical terms that do not refer to sensory experience.

Being a village realist myself, my heart is with Vogel, but I think a complete victory is not yet at hand.

BERENT ENÇ

University of Wisconsin/Madison

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NOTES AND NEWS

The editors of the Journal are sorry to announce the death of Victory Pomèranz, Business Manager of the Journal, 1962–87. Vicky was hired in 1962 by Leigh Cauman, then Managing Editor, for a three-week project and stayed for twenty-five years, during which time she also became a James Joyce scholar. She was named for Armistice Day, her birthday. We will miss her.

The American Society for Eighteenth-Century Studies announces the 1989–1990 James L. Clifford Prize which goes to the best nominated article, an outstanding study of some aspect of eighteenth-century culture, interesting to any eighteenth-century specialist, regardless of discipline. The article should be no longer than 7500 words and must have appeared in print in a journal, festschrift, or other serial publication between July 1989 and June 1990. The article may be nominated by a member of the society, or by its author. Nominations must be accompanied by an offprint or copy of the article and must be postmarked by February 1, 1991. The winning author must be a member of the Society at the time of the award. Send nominations and inquiries to ASECS, Edward P. Harris, Executive Secretary, Univ. of Cincinnati, Mail Location 368, Cincinnati, OH 45221.

An Eco-Philosophy Symposium is scheduled in Nairobi, Kenya in August 1991. For information, you may write to Hwa Yol Jung, Moravian College, Bethlehem, PA 18018.

Harvard Law School announces Liberal Arts Fellowships to college and university teachers in the arts and sciences for the academic year 1991–92. The purpose of the fellowships is to enable teachers in the social sciences or humanities to study fundamental techniques, concepts, and aims of law, so that in their teaching and research they will be better able to use legal materials and legal insights which are relevant to their own disciplines. Applications for 1991 should be complete by January 15, 1991. For more information, write to the Chair, Committee on Liberal Arts Fellowships, Harvard Law School, Cambridge, MA 02138.

The London School of Economics and Political Science announces the Lakatos Award in Philosophy of Science for an outstanding contribution to the philosophy of science in the form of a book published in English. The 1991 Award will be for a book published during the last six years (bearing an imprint from 1985 to 1990 inclusive). The closing date for nominations is April 15, 1991. For more information, write to Lakatos Award, London School of Economics, Houghton Street, London WC2A 2AE, England.

The *Mind* Association announces an international conference to commemorate the centenary of the publication of F. H. Bradley's *Appearance & Reality*. The colloquium will be held April 2–5, 1993, at Merton College, Oxford. Initial deadline for submission of papers is October 1991. For more

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

information, write to Guy Stock, Dept. of Philosophy, Univ. of Dundee, Dundee DD1 4HN, Scotland.

The National Academy of Sciences/National Research Council of the U.S. National Committee for the International Union of the History and Philosophy of Science announces travel grants available for the IX International Congress of Logic, Methodology, and Philosophy of Science in Uppsala, Sweden, August 7–14, 1991. Applications for travel grants may be obtained by writing to Milton A. Whitcomb, National Research Council, U.S. National Committee for the International Union of the History and Phil. of Science, 2101 Constitution Avenue (HA 186), Washington, D.C. 20418.

Yale University announces the fall completion of the Yale Plato Microfilm Archive. This project, initiated by Robert Brumbaugh and Rulon Wells in 1957, is a collection on microfilm of all known extant pre-1600 manuscripts of Plato's works, in whole or part. The collection is located in the Yale Libraries and is available for study. Inquiries may be addressed to Heinrich von Staden, Classics Dept., Yale Univ., Box 1697 Yale Station, New Haven, CT 06520, or Robert Brumbaugh, 150 Ridgewood Ave., North Haven, CT 06473.

Barnard College and the Graduate School of Philosophy of Columbia University wish to announce that Hide Ishiguro has left and moved to the Department of Philosophy, Keio University, Mita, Minato-ku, Tokyo, Japan.

The Department of Philosophy at Columbia University is pleased to announce the appointments of Haim Gaifman, formerly of the Hebrew University of Jerusalem, as Professor of Philosophy, and John Collins, formerly of the University of Pennsylvania, as Assistant Professor. Julia Annas, currently of the University of Arizona/Tucson, will join the department as Professor of Philosophy in July 1991.

The Department of Philosophy at the University of Florida is pleased to announce the appointments of John Biro as Professor and Chairman, and Paul Berckmans and Stephen Jacobson as Visiting Assistant Professors.

The Philosophy Department at the University of Miami is pleased to announce the appointments of Susan Haack as Professor, Alan Goldman as Chair, and Risto Hilpinen as Research Professor.

The Department of Philosophy at Valdosta State College is pleased to announce the appointment of Aristotelis Santas as Assistant Professor.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVII, NO. 12, DECEMBER 1990

TRUTH AND LANGUAGE-WORLD CONNECTIONS*

Philosophers have notoriously had little success in their search for a satisfactory, substantive correspondence theory of truth.[1] According to the early Bertrand Russell, true beliefs correspond to facts: the fact that snow is white makes the belief that snow is white true. The difficulties such fact theories face are well-known. For example, none of the explanations of falsehood is close to being satisfactory. One suggestion is that beliefs are false because there are no facts to which they correspond. This kind of theory easily degenerates into the *existence* theory (A. N. Prior's terminology, *op. cit.*) of truth. In the *Theaetetus,* Plato wrestles with the dilemma posed by the existence theory: if sentences are false when there is nothing to which they correspond, then false sentences would appear to be about nothing. Part of the solution to the difficulty, offered in the *Sophist,* is to recognize that sentences have both subjects and predicates; for then a sentence can be about something

* I thank Michael Devitt, Anil Gupta, and Jerry Kapus for their comments on earlier drafts. I am afraid I have probably not adequately addressed all of Devitt's doubts. Thanks also to the philosophers at Victoria University/Wellington, for their helpful comments.

[1] See, for example, A. N. Prior's survey of the correspondence theory in "Correspondence Theory of Truth," in *Encyclopedia of Philosophy,* Paul Edwards, ed. (New York: Macmillan, 1967). By a *substantive correspondence theory of truth* I mean a theory that construes truth as a substantive relational property connecting language or thought with an extralinguistic reality. I include the correspondence theories of Russell, and I shall include Hartry Field's.

Some philosophers (e.g., Prior) follow Alfred Tarski in claiming that a formally correct definition that has as consequences, so-called T-sentences, like

'Snow is white' is true iff snow is white.

captures all there is to capture in correspondence theories of truth. On this view, nothing further needs to be said about truth "relating" language to the world. Theories like this, and others, have been called *deflationary.* It is expected that deflationists assume truth does not have a causal explanatory role.

0022-362X/90/8712/671–687 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

(Theaetetus), and yet false (Theaetetus flies). This does not leave us within reach of a theory of truth and falsity, however.

Alternative proposals for falsehood have encountered other difficulties. Russell[2] suggests that false beliefs (false sentences, or false propositions) might correspond to objective falsehoods; or, alternatively, that belief is a relation between the believer and an indefinite number of objects. Then later Russell[3] proposes that a true sentence corresponds in one way to a fact and a false sentence corresponds in a different way to the same fact. What are facts, and what are objective falsehoods? What kinds of language-world connections are there that might provide the assumed correspondence?

The picture theory of the *Tractatus* might be viewed as offering a natural explanation of correspondence—for one is supposed just to *see* what a picture pictures. But it is pretty obvious we do not just see what a picture pictures, for different people with different backgrounds see many different things.[4] Furthermore, how can a mere configuration of natural linguistic objects (or names) be propositional? To the extent that a picture theory more satisfactorily addresses this question, the picture analogy seems undermined. Relations (of implication, for example) that obtain between sentences do not obtain between pictures. If we loosen up on the picture analogy, what explanation do we have of the connections between language and reality?

Faced with problems such as these, many philosophers have abandoned the idea that the correspondence-to-facts framework will lead to an acceptable theory of truth (or meaning). Indeed, alternatives to a substantive correspondence truth theory have been offered. There is Alfred Tarski's theory, according to which truth definitions that are formally correct and materially adequate capture all that is to be captured in the notion of "correspondence"; there are the disquotational theories,[5] and the prosentential theory of truth.[6]

[2] "On the Nature of Truth and Falsehood" (first published in 1910), in *Philosophical Essays* (New York: Simon & Schuster, 1966).

[3] "The Philosophy of Logical Atomism," reprinted in *The Philosophy of Logical Atomism*, David Pears, ed. (LaSalle: Open Court, 1985).

[4] See Nelson Goodman, "The Way the World is," in *Problems and Projects* (New York: Bobbs-Merrill, 1972). Another problem: it has been claimed that pictures picture what they picture in virtue of certain conventions. Can an account be given of the conventions governing verbal "pictures" without assuming language already understood?

[5] See W. V. Quine, *The Philosophy of Logic* (Englewood Cliffs, NJ: Prentice-Hall, 1970), pp. 10–13; Steven Leeds, "Theories of Reference and Truth," *Erkenntnis*, XIII (1978): 111–129; Paul Horwich, "Three Forms of Realism," *Synthese*, LI (1982): 181–201; and Scott Soames "What is a Theory of Truth?" this JOURNAL, LXXXI, 8 (August 1984): 411–429.

[6] See Grover, Joe Camp, and Nuel Belnap, "A Prosentential Theory of Truth,"

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Those seeking a substantive truth theory—usually because they think truth has a causal explanatory role—think, however, that none of these "deflationary" alternatives provides adequate explanations of truth as a needed language-world relational property. Hartry Field[7] is one such philosopher. Convinced that truth has an explanatory role, Field offers a suggestion as to the form a substantive correspondence theory might take. His suggestion is that something along the lines of the causal-historical theories of reference should be employed to provide an account of truth as a substantive relational property connecting language with the world. Because we need a better understanding of the possible points of contact between *truth* and *language-world connections,* it will be useful to examine this provocative suggestion of Field's. Field's paper is important because of the deep issues it raises; for that reason, it has also been controversial.[8] Despite this attention by philosophers expressing widely divergent views, an interesting feature of Field's suggestion seems to have gone unnoticed: it does not accomplish what I believe Field (and, I presume, many readers) thought it would accomplish—it does not lead us to a theory in which truth would have a *distinctive* substantive role in connecting language to the world.

My main goal is to highlight, in section IV, this curious feature of a correspondence theory of truth that incorporates as its substantive language-world connections the language-world connections of the causal-historical theory of reference. As a step toward this end, I first outline, in section I, Field's assessment of Tarski's truth definition. In section II, taking up Field's suggestions, I sketch a few of the details of how the causal-historical theory of reference connections might be incorporated in a Tarskian theory of truth. In Section III, I

Philosophical Studies, XXVII (1975): 72–124; my "Inheritors and Paradox," this JOURNAL, LXXIV, 10 (October 1977): 590–604; my "Truth: Do We Need It?" *Philosophical Studies,* XL (1981): 69–103; and my "On Two Deflationary Theories of Truth," in *Truth and Consequences,* M. Dunn and A. Gupta, eds. (Boston: Kluwer, forthcoming); R. Brandom, "Reference Explained Away," this JOURNAL, LXXXI, 9 (September 1984): 469–492; and K. Wilson, "Some Reflections on the Prosentential Theory of Truth," in *Truth and Consequences.*

[7] "Tarski's Theory of Truth," this JOURNAL, LXIX, 13 (July 13, 1972): 347–375.

[8] For example, Devitt [*Designation* (New York: Columbia, 1981), and *Realism and Truth* (Princeton: University Press, 1984)] and Michael Friedman ["Truth and Confirmation," this JOURNAL, LXXVI, 7 (July 1979): 361–382] have endorsed Field's claim that something like causal-historical explanations of reference may constitute a crucial element of a theory of truth; whereas Donald Davidson ["Reality without Reference," *Dialectica,* XXXI (1977): 247–258, Hilary Putnam [*Reason, Truth and History* (New York: Cambridge, 1981), ch. 2], John McDowell ["Physicalism and Primitive Denotation: Field on Tarski," *Erkenntnis,* XIII (1978): 131–152], Leeds (*op. cit.*), Robert Stalnaker [*Inquiry* (Cambridge: MIT, 1984), ch. 2], and Soames (*op. cit.*) have raised some doubts.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

briefly reflect on a criticism of Field's program which is pertinent to the issue I pursue. Then in Section IV, after defining falsity, I show that true sentences are not uniquely connected to the world on Field's proposal. In the remainder of the paper, I argue that language-world connections, like those of the causal-historical theory of reference, fail to provide truth itself with the explanatory role Field speculates truth has. I conclude with the suggestion that it seems we can, and perhaps should, treat the issues of language-world connections, and truth and falsity, separately.

I. FIELD'S PROGRAM

In his 1972 paper criticizing Tarski's truth definition, Field proposed that Tarski's theory needs to be supplemented because truth has an explanatory role. For according to Field, Tarski's definition does not accomplish the goal Tarski set himself—a definition that would satisfy the demands of a physicalist who believes in the unity of science. He says, further, that Tarski's definition is only language-specific. Field argues that, if Tarski's definition of truth were modified so that truth for arbitrary languages would be defined, then 'denotes' and 'applies' would appear in Tarski's base clauses; so truth is, in effect, explained by Tarski only with the help of other semantic terms. Field proposes eliminating these terms in a way that would satisfy physicalists.

Thus, the recursive clauses of Tarski's truth definition provide (only) the first important step of Field's definition. For Field's Tarskian base clauses for primitive terms and atomic sentences contain the semantic terms 'denotes' and 'applies'. Field suggests that the causal-historical theory of reference (or something similar) be used to explain 'denotes' and 'applies'. Let us take a close look at the suggestion that the causal-historical theory of reference might yield the kind of truth theory Field seeks. First, 'true' would be eliminated in favor of 'denotes' and 'applies'; then 'denotes' and 'applies', in their turn, are to be eliminated with the help of recent causal-historical theories of reference.

This discussion will move more smoothly if I assume, as Field would need to, that causal-historical theories provide an analysis of reference that would satisfy physicalists. And, despite my preference for another theory of truth, I shall also for the moment follow Field in assuming that truth is a relational property, and in particular, a property of sentence tokens.

II. FIELD'S TRUTH DEFINITION

As a first step toward a truth definition which will be applicable to all languages, and from which all semantic terms can eventually be elim-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

inated, Field offers the following base clauses for selected primitive terms and atomic sentences. (I ignore his function symbols. As Tarski's recursive clauses are not in dispute, I ignore those also.) I follow Field in using '. . . is true on *s*' for a given sequence *s*, rather than '*s* satisfies . . .'.

For a given sequence *s*,

A. (i) the *k*th variable denotes the *k*th member of *s*;
   (ii) the *k*th constant denotes what it denotes.

B. For a given predicate '*F*' and term *e*, '*Fe*' is true if and only if
   (i) there is a (unique?) object *x* that '*e*' denotes; and
   (ii) '*F*' applies to *x*.

These clauses provide only a first step toward a definition; for 'denotes' and 'applies' have yet to be eliminated. Beyond directing us to Kripke's causal-historical account of reference, Field does not give us details as to how the elimination is to be achieved.[9] Also, note that the above definition yields a truth definition for sentence types rather than tokens. As I mentioned above, Field intends a definition for sentence tokens, in part because he says he wants to provide a definition that will cope with ambiguity. I shall usually ignore this detail.

According to causal-historical theories of reference, tokens of singular terms are causally[10] connected to objects, and tokens of predicates like count nouns, natural-kind terms, etc., are causally connected to the members of sets of objects. Unfortunately, these details do not suggest unique ways of eliminating 'denotes' from Field's base clauses, even in outline. So I shall provide only schematic elimination clauses.

For a given sequence *s*, 'denotes' can be eliminated from the A-clauses and B(i) by means of

D. A token *x* *denotes* *y* iff *x* is appropriately causally connected to *y*.

The situation for predicates [which is what we must look to in unpacking 'applies' in B(ii)] is more complicated.[11] Whereas for a singular-term token only one extralinguistic object is usually as-

[9] This is not surprising, for Saul Kripke's "Naming and Necessity" [in *Semantics of Natural Language*, D. Davidson & G. Harman, eds. (Boston: Reidel, 1972), pp. 253–355] had not at that time appeared in print.

[10] I shall often speak of 'causal connections' rather than the longer 'causal-historical connections'.

[11] Field refers us to Kripke, *op. cit.* It happens that all of Kripke's examples involving 'water', 'gold', etc., have these terms occurring in subject place. I do not know whether this is deliberate. I shall follow Field in assuming Kripke intended his account of reference to apply also when such terms occur in predicate position.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

sumed involved, for a predicate "baptisms" typically involve a set of sample objects, a subset of which are paradigmatic samples of the predicate's extension. Bits of junk are likely to be included along with the paradigmatic samples in the initial baptismal sets. In the case of 'water' and 'gold', for example, the samples may include dirty or salty water, and fool's gold. "Experts" in the society shoulder the responsibility of doing some sorting out here, for facts about the world determine extension; and the experts are supposed to know about the world. They must sort out which or what among baptismal samples are paradigmatic to a determination of extension, as well as the complete extension of the predicate. I cannot offer ways of handling (B) [or (A)] which will do justice to the details of causal-historical theories of reference. Appropriate eliminations must presumably rely, however, on something like the following elimination schema, for a given sequence *s*,

AP. A token '*F*' *applies* to *x* iff

(a) there is a (possibly empty) paradigmatic set of objects *f'* included in a baptismal set of sample objects; and the members of *f'* are appropriately causally connected to the token '*F*';[12] and

(b) there is a (possibly empty) set *f*—the extension of the token of '*F*'—such that the members of *f* are similar in appropriate respects to the members of *f'*.

Tarski allows primitive predicates with empty extensions. As Field does not comment on this point, I assume he also allows empty extensions for predicates. I do not know how he would do this. Nevertheless, I have allowed that the set of paradigmatic samples, *f'*, included in the baptismal set, could be empty; and I have allowed that the extension of the token of '*F*' could be empty. [Perhaps definite descriptions are used to pick out the members of the initial *f'* when *f* (and *f'*) are empty; and sometimes nothing satisfies the definite description. Or perhaps *f'* could be empty in another way: if causal connections are involved, the baptismal set will have to be nonempty; however, *f'* would be empty if the baptismal set contained no objects that can be counted as paradigmatic.]

AP(a) and (b) do not alone suffice to eliminate 'applies', of course. They ensure only that the terms in a token of *Fe* have *extensions*. Given the syntactic structure of *Fx*, I complete AP with

(c) *x* is a member of *f*.

Note that I could just as easily have used '*x* is *F*' in place of '*x* is a member of *f*'. Either would give the right extension for 'true', *pro-*

[12] I assume the causal connection is not with the set *f'*, but with its members.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

viding the other clauses do their job. For we usually explain set membership to those new to set theory by means of more commonplace predicates, e.g., we explain 'Adam is a member of the set of fathers' by means of 'Adam is a father'.[13]

This concludes my schematic development of Field's proposal for a causal-historical truth definition.

III. CRITICISM OF FIELD'S PROPOSAL

Field's proposal for a substantive correspondence theory seems, in some respects, an advance on earlier proposals for a substantive theory. For Field does not assume that there is a correspondence without making some attempt at identifying the "correspondence"; and, following Tarski, he does not appear to assume connections with facts.

Despite these positive features, there have been criticisms of Field's proposal, as I mentioned above. Pertinent to the point I want to make is an issue that relates to a criticism John McDowell raises. McDowell questions Field's location of language-world connections at the level of primitive reference. Field says that language-world connections at the level of sentences are to be explained or derived from language-world connections at the level of the primitive terms of the language—via Tarski's recursive clauses. The language-world connections at the level of primitive terms on which I am focusing are those of causal-historical theories of reference. Let us consider, then, the task of drawing on the details of causal-historical theory of reference as we try to fill in the gaps in the schematic elimination clauses D and AP for 'denotes' and 'applies'.

We have been told that referential chains are determined by complex social linguistic factors. Saul Kripke and Hilary Putnam[14] give us some of these details. For example, scientific experts have a role in the determination of reference, etc. Charles Chastain[15] provides other kinds of details; for example, he illustrates some of the ways in which proper names can be introduced into a discourse:

D14. A man named Fred Schultz runs a liquor store around the corner from here. He was held up yesterday, and . . .

D15. A man known to the underworld as Joe the Snake was found dead in a vacant lot this morning. Police spokesmen stated . . .

[13] Given our lack of understanding of the set-theoretic paradoxes, it is doubtful we should view 'x is a member of f' as contributing to a better explanation than 'x is an F' would.

[14] "The Meaning of 'Meaning'," in *Language, Mind, and Knowledge*, K. Gunderson, ed. (Minneapolis: Minnesota UP, 1975), pp. 131–193.

[15] "Reference and Context," in *Language, Mind, and Knowledge*, pp. 194–269.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

D17. A certain bookie I know whom I'll just call "Harry" told me that . . .

D18. Stella Houston (a fictitious name) underwent a sex-change operation six months ago. "She" told this reporter that . . . (*ibid.*, pp. 216–7).

Of these Chastain says:

> Sometimes it appears that a proper name is introduced *de novo* into a context, in such a way that it is not referentially linked to another occurrence of the same name in another context. This is indicated explicitly in D17 by the phrase 'whom I'll just call "Harry" '; the pseudonym is introduced simply in order to have a name *in that context* for the person whom the speaker is referring to by another singular term ('a certain bookie I know'), and previous occurrences of the name in other contexts are irrelevant. . . . Sometimes the situation is just the opposite: the first occurrence of the name in the context harks back to previous occurrences in other contexts, as in . . . D14, D15 . . . (*ibid.*).

The causal-historical explanation of reference portrays a situation where complex human linguistic behavior is causally or historically linked to extralinguistic reality. Our D and AP clauses must take account of the fact, for example, that both the immediate discourse, and earlier discourses—as well as the expertise, intentions, and theories of speakers of the language—have a role in determining reference. In particular, our elimination clauses must take account of the fact that initial baptisms typically involve sentences, not just terms. Putnam and Kripke illustrate this aspect of their theories with examples like 'This (liquid) is water', 'Gold is the substance instantiated by the items over there, or at any rate, by almost all of them'. Whole sentences (not just terms) also typically appear at later stages of referential chains, as in D14 and D15, for example.

These excerpts serve to illustrate partially the criticism that the causal-historical theory of reference does not provide Field with what his program would need. For the causal-historical theory does not have language-world connections occurring first at the level of primitive terms, and then only derivatively at the level of sentences —as Field's project seems to assume. Field needs to do some sorting out here.[16]

The point I want to draw attention to in all of this is that causal-historical theories do not assume *true* sentences have a unique role in

[16] Devitt thinks there are optional strategies here; e.g., along the lines suggested in *Designation*, pp. 113–124; and in *Realism and Truth*, ch. 10.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

causal-historical chains—false sentences, as well as true sentences, appear in the chains.

This observation raises an initial question concerning Field's program. Correspondence theorists usually assume that the language-world connections of true sentences should be different from the language-world connections of false sentences. Since true sentences do not appear to have a unique role in causal-historical theories of reference, one wonders whether a truth theory of the kind Field proposes will yield different language-world connections for true and false sentences. This is the issue I am chasing.

IV. FALSEHOOD

Field's definition yields extensions for only 'true' and, by implication, 'not true'. 'False' is left to us. A question arises when we think about how we should handle the existential quantifiers in the above clauses. Both Field's A-clauses and his overall discussion suggest that he follows Tarski in assuming that the languages under consideration contain only denoting singular terms. As this is the simplest assumption to accommodate, I shall proceed on that assumption in defining falsity. So the base clauses, A(i) and A(ii), will be the same for falsity as for truth.

It remains to define 'A token of *Fe* is false'. For this, I shall construct C-clauses, following as closely as possible the above B-clauses. C-clause can be the same as B-clause up to the end of (i), since we continue with the assumption that singular terms denote. As I am defining falsity and not 'not true', I cannot simply add '*F* does not apply to *x*', since I assume Field would hold that all predicates appearing in the primitive sentences of his languages have extensions—though possibly empty. For this reason, C(ii) will be more complex than B(ii). The following should suffice.

For a given sequence *s*,

C. A token of '*Fe*' is *false* on *s* if and only if
  (i) There is a (unique?) object *x* that the token '*e*' denotes; and
  (ii) (a) The given token of '*F*' has a (possibly empty) extension; and
    (b) '*F*' does not apply to *x*.

D and AP, above, can be used to eliminate 'denotes' and 'applies'. That leaves 'has an extension', which for a given predicate is assumed to be causal-historically determined. Note that AP(a) and (b) contain all we need for an elimination clause. Rewritten, in the form of a (schematic) definition we get for a given *s*,

E. A token of '*F*' has an *extension* iff
  (i) There is a (possibly empty) paradigmatic set of objects *f'* included

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

in a baptismal set of sample objects; and the members of f' are appropriately causally connected to the token of F; and

(ii) There is a (possibly empty) set f such that the members of f are similar in appropriate respects (as identified by the experts) to the members of f'.

Providing the details could be filled out, this would yield a definition of falsity from which all semantic terms would be eliminated. In general, 'F' does not apply to x if any of AP(a), (b), or (c) were denied. On the assumption that 'F' has an extension [i.e., C(ii) (a) obtains], however, 'Fe' will be false only if x is not a member of f.

In summary, we have:

'Fe' *is true,* on s, just in case,
e is appropriately causally connected to x (its denotation); 'F' is appropriately causally connected to certain members (the objects in f') of f ('F's extension); and x is a member of f;
'Fe' *is false,* on s, just in case,
e is appropriately causally connected to x (its denotation); 'F' is appropriately causally connected to certain members (the objects in f') of f ('F's extension); and x is not a member of f.

With the base clauses of the definition set out in this fairly tidy way, we can see immediately where the causal-historical theory establishes substantive causal-historical connections between sentences and the world. The connections are between some of the parts of a sentence and extralinguistic objects. We also see that, since a sentence contains the same words whether or not it is true or false, these causal-historical connections are the same for a sentence whether or not it is true or false.

It is only in the final clauses of the definitions for " 'Fe' is true" and " 'Fe' is false" that there is a difference. In the case where 'Fe' is true, it is required that x (the object which is causally connected to e) belong to f (the causally determined extension of 'F'); in the case where 'Fe' is false, it is required that x not belong to f. This is a difference that has nothing to do with the causal historical language-world relations, however.[17] It concerns only extralinguistic reality—if one is a realist. Indeed, this difference between true and false sentences is one that deflationists recognize. Consider Tarski's requirement of material adequacy, according to which a theory of truth should have among its theorems T-sentences like " 'Snow is white' is true if and only if snow is white."

[17] Field's use of 'applies' makes it easy for us to overlook this point.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

And so my point can now be made:

> The causal-historical connections between language and the world—which have been incorporated in this proposed definition of truth—are the same for a sentence whether or not the sentence is true or false.[18]

This feature of Field's language-world connections is perhaps not surprising, when one considers that the language-world connections Field utilizes come from causal-historical theories of reference. According to these theories, the same kind of story is to be told about 'this ring' and 'gold', for example, whether or not tokens of these terms appear in declaratives like 'This ring is gold' and 'This ring is not gold', or in questions like 'Where is the gold ring?' and 'Whose is this ring?', or in commands like 'Give me that ring!'. No presently proposed theory (to my knowledge) suggests that different kinds of causal-historical referential connections obtain depending on whether a term occurs in a question, a command, or a declarative—or in a true declarative or a false declarative.

It is true that Field is cautious in suggesting that the causal-historical theory of reference provides correspondence theorists with exactly what is needed.

> [Kripke] has suggested a kind of factor involved in denotation that gives new hope to explaining the connection between language and the things it is about. It seems to me that the possibility of *some such* theory of denotation (to be deliberately very vague) is essential to the joint acceptability of physicalism and the semantic term 'denotes' . . . (*op. cit.*, p. 367).

The vagueness makes it hard to tell, however, what 'some such theory' includes. If Field includes theories that are motivated by issues that we traditionally associate with "meaning," then we should be aware that the same substantive language-world relations might emerge for a sentence, whether it is true or false.

Consider for a moment, and more generally, some of the factors that can influence the selection of language-world connections in a theory that seeks a replacement for, or an account of, denotation. If we suppose that there are certain language-world connections that determine (or partly determine) the significance of a word in a sen-

[18] Like the early theorists, Field begins with an idea that sounds appealing initially—when one thinks only about truth. As soon as one tries to complete the theory with an account of falsity, however, the theory appears to raise more (important) questions than it answers.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tence, these language-world connections must presumably be the same whenever the significance of the word remains the same. Given, further, that the contribution the component words make to a sentence is usually more or less fixed, we assume that a sentence has roughly the same meaning, whether it is true or false.[19] (I ignore the interesting complications that indexicals and vagueness introduce.) Accordingly, we can expect that the kinds of language-world connections to which Field directs us might be the same, whether the sentences in question are true or false.[20]

There are a number of reasons for thinking Field may have hoped for different language-world connections for true and false sentences, however. One reason is that he believes truth has an explanatory role. I discuss this issue in the next section. There is also the following.

> . . . *Part* of the explanation of the truth of 'Schnee ist weiss und Gras ist grun', presumably, would be that snow is white and grass is green. But this would only be part of the explanation, for still missing is the connection between snow being white and grass being green on the one hand, and the German utterance being true on the other hand (*op. cit.* pp. 359–360).

In this passage, Field tells us what he is looking for by way of substantive language-world connections. Part of the requirement is captured in AP(c)—the deflationary Tarski part. Beyond that, he seeks a connection between the sentence and 'snow being white and grass being green'. Field's way of stating this requirement unfortunately brings to mind the problems the earlier fact theories encountered. For consider the sentence, 'The cat is on the mat', when the cat is up the tree. Is the sentence connected to the cats not being on the mat (a negative fact), or the dogs being on the mat, or the cats being up the tree?[21] What we do know is that 'The cat is on the mat', if false, cannot be connected to the cats being on the mat, because

[19] We make this assumption when we judge that sentences of the form '*S* & not *S*' are contradictory.

[20] I have shown, in sect. III, that true sentences do not have a special role in the version of the causal-historical theory of reference I considered. From this, I cannot conclude there is no theory of denotation according to which some true sentences (and no false sentences) are uniquely connected to the world. For what it is worth, however, I am almost certain there are no interesting causal connections that obtain between all and only true-sentence tokens and extralinguistic reality.

[21] These questions are taken from Peter Strawson, "Truth," *Proceedings of the Aristotelian Society*, Supp. Vol. XXIV (1950): 129–156.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

there is no such thing—the language-world connections must presumably be different for true and false sentences, if we approach language-world connections from this direction.

I believe, however, that Field's talk of 'snow being white', etc., is only a manner of speaking that we should not take seriously. Let us back off a bit. The crucial test for Field's theory will come in those places where he claims truth has an explanatory role. Does Field *need* a theory of truth according to which there are different language-world connections for true sentences and false sentences?

V. A CAUSAL EXPLANATORY ROLE FOR TRUTH

A reason Field gives for advocating a substantive correspondence theory of truth is that he thinks truth has a causal explanatory role, otherwise he seems to think a disquotational theory of truth would do. His most recent defense of these claims appears in "The Deflationary Conception of Truth."[22] In that paper, Field speculates that certain kinds of systematic reliability, T-reliability, can only be explained with a "full-blooded" correspondence theory. The task is

> . . . the matter of explaining T-reliability (as opposed to using T-reliability in explanations). If, under a given interpretation J, an agent's states are highly T-reliable and converse T-reliable—i.e. for certain sorts of p, when the agent has a belief whose truth-conditions under J are that p, then generally p; and conversely when p then the agent has a belief whose truth-conditions under J are that p—then this fact is a striking regularity that calls for some sort of explanation. Just what the explanation will be like will of course vary greatly from case to case, depending on the interpretation J and the nature of the class of p for which the agent is T-reliable (and converse T-reliable) under J. Suppose we have a successful social activist, whose success is explained in part by her extensive knowledge of social movements of the past. The interpretation useful in explaining her success will make her T-reliable (and converse T-reliable) about such events as the Kronstadt rebellion; and we will need an explanation of the striking correlation between her belief-states and the events involving the Kronstadt sailors. It seems likely the explanation would have to involve something like causal networks of information . . .
>
> A familiar proposal for giving a correspondence theory of truth (perhaps most explicit in Field, 1972) is to explain the truth conditions of sentences in terms of reference or reference-like relations for the basic parts of the sentence, and then to give explanations of the reference-like relations for the parts; and it is often assumed that something like causal

[22] In *Fact, Science, and Morality: Essays on A. J. Ayer's Language, Truth, and Logic*, G. MacDonald & C. Wright eds. (New York: Blackwell, 1986), pp. 55–117.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

networks of information of the sort just alluded to will play a role in the theories of reference for the parts. Abstracting from the details of the proposal, it seems that a central feature of the proposal is to build into the theoretical explication of correspondence truth just the sort of thing that is needed to *explain* the striking T-reliability and converse T-reliability that agents have (about some matters), under the interpretations that we find useful in explaining their success. . . . And perhaps it is not surprising that the explanation of such a systematic T-reliability would have to involve the systematic elements required of an explication of the notion of truth conditions (*ibid.*, p. 104).

Let us suppose Field is right when he suggests that causal connections will be a "central feature" of explanations of T-reliability. Has Field built into his "theoretical explication of correspondence truth just the sort of thing that is needed to explain . . . T-reliability"?

I assume one thing we should require of a good explanation of T-reliability is that it provide an explanation of why some people are T-reliable and why others are not T-reliable. So it would seem to follow that, if certain kinds of causal connections between language or an observer and the world are what is needed to explain T-reliability, then the causal connections that obtain when a person is T-reliable will not obtain when a person is not T-reliable. Otherwise, the presence of the causal connections will contribute little (if anything) to an explanation of the differences between being T-reliable and failing to be T-reliable.

If we continue assuming the causal-historical connections of current theories of reference, then the utterances of articulate coherent speakers will tend to be appropriately causally connected to the objects about which they talk. Articulate coherent speakers can have false beliefs, however, and so fail to be T-reliable. Indeed, the presence of referential causal-historical connections—though they may have a role in explaining articulateness—seem to have little to do with a person's T-reliability or lack of T-reliability. For reasons such as these, I conclude that the language-world connections Field suggests we consider incorporating in his truth theory are unlikely to help us distinguish between those articulate persons who are T-reliable and those articulate persons who are not T-reliable.[23]

A good explanation of T-reliability, along the lines Field seeks, would seem to call for different language-world connections for true

[23] Friedman, *op. cit.*, argues that a truth theory like Field's is needed to justify scientific methods of confirmation. I believe that such a truth theory is ill-suited to that task also.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

and false beliefs.[24] As Field's proposal for a theory of truth does not provide the kind of truth property he needs, we face a problem.

What can be salvaged? We have considered the possibility of incorporating in a theory of truth the language-world connections of reference. If truth is to have a causal explanatory role, however, then it seems there must be different causal connections with the world for true and false sentences. And this, we have seen, seems to make the causal connections of reference unsuitable for the task. If truth has an explanatory role we need to look somewhere else for the required language-world connections.

What are our options if truth does not have an explanatory role? While I am inclined to drop the idea of incorporating language-world in our analysis of truth, many philosophers find the spirit of Field's project appealing.[25] Would there be any harm in incorporating something like the language-world connections of reference in an account of truth?

One such approach might proceed as follows. We are free to stipulate how we will use 'true', 'false', 'truth', and 'falsity'; and so, as it pleases us, it seems we could use these terms to designate an identifiable set of causal relations. The causal-historical connections of theories of reference seem relevant to an explanation of linguistic activity. They seem pertinent, for example, to an explanation of meaning, or to an explanation of the fact that a given sequence of sounds is a token of a sentence in a given language. Incorporation of referential connections in an analysis of truth would then seem an option, since harmless. In the particular case we are considering, if referential causal-historical connections have an explanatory role (in explaining meaning, perhaps) then truth would also—though admittedly this would not give truth and falsity explanatory roles of their own.

[24] I suspect connections like those often associated with justified belief are more likely to be relevant.

[25] I believe a reason the incorporation of language-world connections is initially attractive—whether or not one has a view on explanatory role—is that this move addresses a problem for truth *created* by correspondence theorists. Their move to a metalinguistic truth predicate, as in " 'Snow is white' is true," leads one to ask how it is that the sequence of marks on paper which constitute the inscription 'Snow is white' come to be used to say what we say when using it. If we think, instead, of 'true' as functioning at the level of the object language—as 'she' and 'he' do—this question does not arise in connection with truth and falsity. Of course, it does arise when we start thinking about language, no matter what theory of the truth predicate we adopt.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Those who endorse deflationists' accounts of the utility of the truth predicate[26] have reasons for not endorsing the above supposedly harmless suggestion.[27] I shall explain briefly.

According to some who advocate either a disquotational or prosentential theory, 'true' and 'false' are only useful words; since they only provide us with a way of generalizing with respect to sentences, a way of achieving linguistic ascent and descent, etc. That is, truth and falsity, and 'true' and 'false', rather than provide or contain answers to questions regarding such issues as language-world connections, T-reliability, meaning, and realism, provide only the expressibility that we need to say some of the things we want to say, and to ask some of the questions we want to ask about these and other issues. For example,

(i) Is there a causal relation that obtains between a sentence and the world when it is true, but does not obtain when it is false?

(ii) What justifies us in claiming that statements about theoretical entities are true?

These questions do not need to be interpreted as giving truth a property-ascribing role. On the prosentential theory, (i) and (ii) have instances like

(i') Is there a causal relation that holds between 'Reagan cannot remember' and the world, when Reagan cannot remember, which does not hold between the sentence and the world, when Reagan remembers?

(ii') What justifies us in claiming that electrons have mass?

If the expressibility offered by 'true' and 'false' can be separated off from these more substantial issues, then there would seem to be advantages in reflecting this fact in our theories. We are not then faced with the difficulty that there seems no clear intuition as to what more there is to truth. And if we separate the issues, we are less likely to confuse them.

There is, I believe, another reason for not incorporating the language-world connections of reference in a prosentential analyses of truth.[28] Incorporation of language-world connections would pre-

[26] I do not have space to give more details of either the disquotational or prosentential theories. For that, the reader is referred to the references given in footnotes 5 and 6.

[27] Although Field grants that 'true' and 'false' are used in the way disquotationalists have suggested, I do not know whether he would endorse one of the deflationary theories if truth had no explanatory role.

[28] I continue to assume there are no substantive language-world connections that

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

sumably give 'true' a property-ascribing role. If we accept the prosentential explanation of the expressibility with which 'true' and 'false' provide us, then the cost of adding to this prosentential role, a property-ascribing role, is inconsistency. Construed only as a modified prosentence, 'This is false' lacks content because it is ungrounded; whereas if 'false' has a property-ascribing role in addition to its prosentence-forming role, 'This is false' embodies an inconsistency.[29]

I have shown that, in his stimulating papers, Field has presented a confused picture as to the kind of truth theory he thinks we need, since, on the one hand, he has proposed a substantive correspondence truth theory according to which true and false sentences are connected by the same kinds of causal connections to reality. On the other hand, his suggestion that truth has an explanatory role shows we would need different (if any) language-world connections for true and false sentences. The question arises whether we might consider incorporating causal language-world connections of denotation in a truth theory, nevertheless. If 'true' and 'false' provide expressibility of the kind disquotational and prosentential theorists have proposed—and truth itself does not have an explanatory role—I believe there are reasons for not doing so.

DOROTHY GROVER

University of Illinois/Chicago

obtain between only true sentences and the world. The whole situation changes if this assumption is false.

[29] See my "Inheritors and Paradox" for details. The need for Tarski-style truth models does not thereby vanish, if we adopt the prosentential theory; Tarski-style truth models are needed so that we can track the logic of prosentences, modified prosentences, and ungrounded prosentences.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

RIGHTS AND RIGHTS VIOLATORS: A NEW APPROACH TO THE NATURE OF RIGHTS*

In today's world, there is much talk about rights: women's rights, minorities' rights, children's rights, animals' rights, etc. There are groups devoted to seeing to it that each possible right is recognized and respected. And some philosophers have taken it upon themselves to defend the very existence of rights. Why are rights seen to be so important?

Let me suggest that much of the importance we attach to rights derives from our desire to think of ourselves as qualitatively different from the "lower" animals. Part of what makes us different, so the story goes, is that we have more rights than they do, and that our rights are more significant than theirs (both in the sense that we have rights that are in themselves more significant than the rights had by other animals, such as our right to liberty; and in the sense that, in the event of there being a conflict between a person and another animal in respect to what seems to be the same right, say the "right to life," it is always the person's right that is supreme).

Perhaps it is this perceived link between the superiority of persons over animals and the superiority of the rights of persons over the rights of animals that has led so many philosophers to pursue the question 'what is a right?' by considering what it is about persons which differentiates them from "mere" animals. An extension of this line of reasoning has then led them to wonder precisely which beings have rights and in virtue of which of their properties is it that they have these rights.

In opposition to this approach, I would like to suggest that consideration of the properties of a particular being in isolation provides insufficient data to determine the rights of that being; and, furthermore, that the nature of rights in general cannot be determined solely by consideration of the characteristics of those who are said to have rights. (Most of the discussion in this paper will be in terms of human rights but, as I hope will be clear, I take it that what I have to say can easily be extended to talk of any kind of rights—natural rights, legal rights, etc.)

One who holds that a being's rights can be determined by consideration of that being alone might well begin by claiming, in somewhat

* I would like to thank Norman Care, Dan Brock, and Bernard Gert for their helpful comments and, even more, for the encouragement they gave me while working on this paper.

0022-362X/90/8712/688–694 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

more precise language, that sentient beings such as ourselves have the right not to be caused pain, and to add that such beings have this right in virtue of something having to do with their sentience. It is a commonplace, however, to note that broadly phrased rights claims such as this will not do. The dentist to whom I go to have a cavity filled does not violate a right of mine simply because her drilling causes me pain. Nor does the cavity, which caused the toothache, violate my right.

My proposal is that we consider a new question: In virtue of what do we say that a being is capable of *violating* (*or respecting*) a right? It may seem strange to consider what can violate, or "break," a right in order to discover something about rights. Certainly we do not discover much about china vases by considering small children and bulls. Nevertheless, I believe we will find that this question sheds as much light on the nature of rights as does the seemingly more appropriate question: In virtue of what does a being have a right?

To begin our discussion, let us imagine the following scenario: Allison, a fully competent adult person, is sitting at home, quietly reading in her favorite armchair. Suddenly, a crash breaks the stillness of the evening, and Bertha appears in front of her. Bertha leaps at Allison and violently attacks her. Just as suddenly, Bertha disappears, leaving Allison with a broken arm, and covered with scratches and bruises. Can we say of Allison that her rights have been violated?

No doubt it is tempting to say that the above paragraph clearly describes a situation in which a person's rights have been violated. It must be recognized, however, that this is tempting only because we have made certain assumptions about Bertha. But, in fact, we have been told nothing about her. She may be something more or less in keeping with our assumptions, say, a hostile youth who wantonly beats unsuspecting persons. But perhaps she is a tormented paranoid schizophrenic, mistakenly released from a mental institution, or a wild bear that has wandered into Allison's home from a nearby forest.

Most will agree that, if Bertha is a hostile youth, then Allison's rights have been violated, whereas if Bertha is a wild bear, they have not. There will be more disagreement, I believe, over the case in which Allison has been attacked by Bertha the schizophrenic.[1]

[1] There are at least two factors that complicate the case of Bertha the paranoid schizophrenic. First, we are not sure how to assess the rational capabilities of a paranoid schizophrenic: Are we to evaluate his actions on the basis of his beliefs, in which case actions such as those of Bertha may be justified? Or are we to evaluate his actions on the basis of the "real" situation, in which case such actions are often

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

These various evaluations can only be correct if it is true that, in determining whether or not a certain act constitutes the violation of a right, we in fact do take into account not only the characteristics of that which is said to hold the right, but also the characteristics of that which performs the action said to violate the right.

Actions, the external appearances of which are identical, will or will not be said to violate a right depending on whether or not a *person*[2] is performing the action. So, a human-looking robot that malfunctions and stabs a person *would not* be said to have violated a right, whereas a person going through precisely the same motions *would* be said to be violating a right. (A robot that is programmed to stab a person probably is a *means* toward violating that person's right; nevertheless, it is not the robot that violates the right.) Or, to make the same point from the opposite direction, the animals and objects of children's stories can violate rights precisely because they are to be considered as persons. A malicious tea kettle that manages to spill boiling water on someone will be said to have violated that person's right not to be hurt. A real tea kettle that does what appears to be exactly the same thing will not be said to have violated anyone's right.

I would now like to suggest that the criteria that an entity must meet in order to be said to violate another's rights are more stringent than the criteria that an entity must meet in order to be said to have rights. To see this, consider the scene between Allison and Bertha in reverse—in this case Allison unprovokedly attacks Bertha. Most would agree that under *any* of the three descriptions of Bertha it would be said that Allison has violated her rights; and there is no doubt that, under either of the first two descriptions of Bertha (as a hostile youth, or as a paranoid schizophrenic), her rights will be said to have been violated. To take another example: if a small child, in the throes of a temper tantrum, injures an adult, we will not say that the child has violated the adult's rights. If, however, an adult in the throes of a tantrum inflicts comparable injuries on a small child, most of us will say that the child's rights *have* been violated.

Let us stop to consider the case in which a child harms an adult.

---

unjustified. Is a paranoid schizophrenic unable to recognize and respect the rights of others? If he is unable to recognize and respect the rights of others he cannot have violated Allison's right.

The second factor to take into account is that others may have duties concerning Bertha the schizophrenic such that *they* will be said to have violated Allison's right by releasing Bertha.

[2] Although I am not going to attempt to define 'person', it should be noted that I do not intend it to be understood as synonymous with 'human being'.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Let us imagine that she bites him. If the child is a newborn, we would certainly not say that she has violated a right of the adult. Nor are we likely to say that the adult's right has been violated by a toddler in her testing stage. By the time the child is five, however, many will say that she violates the adult's right when she bites him. And by the age of eight or ten almost all would agree that the child does violate the adult's right.

Whether or not one contends that human beings grow into their rights, we will all agree that we grow into our ability to recognize and respect the rights of others (and therefore that we grow into the ability to violate them as well). The newborn is presumably unable to differentiate between her own actions and the actions of those around her, and so it would be inappropriate to hold her responsible for her actions (if she can even be properly said to act). It is probably also inappropriate to hold a toddler to be morally responsible for her actions, but at this age the child is aware of being the originator of her own actions, and so it is now appropriate to attempt to affect her behavior through some sort of discipline. By the age of five, we are pretty sure that the child will have learned that being bitten hurts, and that her biting is therefore unacceptable. And by the time she is ten years old, we are certain that the intellectually and emotionally normal child will have learned that such behavior is forbidden; and we will hold her morally responsible for such actions.

The ability to respect or violate the rights of others thus seems to rely on an intellectual (and emotional?) ability to recognize the harms one can do to another being, and the types of behavior considered to be inappropriate. Some philosophers, such as Michael Tooley,[3] have tried to make the *possession* of a right contingent on a similar level of intellectual competence, but for the most part we recognize no such intellectual prerequisite for the possession of rights. There are many real disagreements in the literature over what is to be said to have rights, and why, and so I shall not attempt to give a list of sufficient conditions for possessing rights. Perhaps all that is common to the various accounts is the idea that the entity must be able to suffer harm (on one or another understanding of 'harm').

The fact that there is so much explicit disagreement over that in virtue of which an entity can be said to have a right, along with the fact that there is such implicit agreement in regard to that in virtue of which an entity can violate or respect a right,[4] leads me to suggest

[3] "A Defense of Abortion and Infanticide," in Joel Feinberg, ed., *The Problem of Abortion* (Belmont, CA: Wadsworth, 1973), p. 60.

[4] As evidenced by the fact that the examples used by those who write on rights never involve inanimate objects violating rights, for example.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that in discussions of rights we are at least as concerned with rights violators or respecters and their behavior as we are with the holders of rights and their possessions.

When we come to see the importance of the concepts of "respecting" and "violating rights" as compared to the concept of "possessing" or "having a right" we can begin to get a clearer picture of both what is correct and what is incorrect about many of the claims often made for rights. For instance, it is often said that the importance of having a right lies in the power it gives one to make a claim for some good. Thus, David Lyons[5] says;

> When A, in particular, holds a certain right against B, A is a claimant against B. A "claimant" is one empowered to press or waive a claim against someone with a corresponding duty or obligation. He can, if he wishes, release the other from his obligation and cancel it, or he can insist upon its performance (*ibid.*, p. 60).

If 'to empower' is taken to mean "to give one the power to . . . ," however, one who has a right is certainly not necessarily empowered to press or waive a claim. A kidnapping victim who is bound and gagged certainly does not have the power to press or waive his claims against his captors. Nor does a comatose patient have this power in regard to her physicians or relatives. Moreover, many say that some rights are inalienable—that is, those who are said to hold such rights are *disabled from* waiving them, such as the rights to life and to liberty. Many persons also feel that at least some animals have some basic rights, such as the right not to be caused pain for no good reason. But to say that the having of a right empowers one to make a claim rules out all animal rights from the very beginning. To this extent, then, the idea that rights empower is incorrect.[6] Similar examples would also show that rights do not empower others to make claims on behalf of the rightholder.

Nevertheless, we do expect others to refrain from kidnapping us and to help us when we are patients and cannot demand such help. To the extent that society is such that it makes these expectations reasonable we may consider ourselves to be thereby empowered. That is, we are empowered in a substantial way by rights only in so far

[5] "Rights, Claimants, and Beneficiaries," in David Lyons, ed., *Rights* (Belmont, CA: Wadsworth, 1979).

[6] This point should not be confused with the assertion that having a right, or being said to have a right, has nothing to do with making claims. The having of a right is almost certainly necessary in order to press or waive a claim, but it is certainly not sufficient; nor does the pressing of a right guarantee that others will act in the desired manner.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

as we are empowered by that behavior of others which we call "respecting rights."

Joel Feinberg[7] adds an interesting twist to the idea that rights are the grounds of claims. He holds that to have a right is to have a *valid claim,* and that to have a valid claim involves the ability of those to whom the claim is made to fulfill that claim. In other words, *A* can only be said to have a valid claim against *B* to do *X* if it is within *B*'s power to do *X*: and *A* can only be said to have a right to *X* if there is someone who is able to do *X*.

An implication of Feinberg's view is that we say of a person that he has a given right only in those situations which involve another entity which is capable of acting in such a way as to be said to be either respecting or violating that right. Therefore, one must know something about the potential rights violator before one can know about a right.

Furthermore, as Feinberg himself notes, we do sometimes want to say of a given person or group of persons that they have certain rights even when, right now, no one is in a position to grant that to which they are said to have a right. He suggests that we admit the existence of claims in such instances, but due to his definition of 'valid claim' he cannot call these claims valid, and therefore cannot associate them with *existing* rights (though he does call them the "grounds" for future rights). This type of reasoning forces Feinberg to say such things as that persons in underdeveloped countries may not have the right to even a minimal amount of food—e.g., if there is no one in a position to provide that food. He leaves it open how one is to justify the push that is often needed to make a "claim," such as the claim to food, into a "valid claim"; that is, how we would justify insisting that we work toward changing the situation so as to allow others, in the future, to be able to fulfill the claim.

In a similar vein, it is said that rights are necessary for a secure society. Richard Wasserstrom[8] tells us that,

> To live in a society in which there are rights and in which rights are generally respected is to live in a society in which the social environment has been made appreciably more predictable and secure. It is to be able to count on receiving and enjoying more objects of value. Rights have, therefore, an obvious psychological, as well as moral, dimension and significance (*ibid.*, p. 49).

[7] "The Nature and Value of Rights," in *Rights,* pp. 46–56.
[8] Richard Wasserstrom, "Rights, Human Rights, and Racial Discrimination," in *Rights,* pp. 78–91.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Rights are often invoked in this way, but does Wasserstrom's conclusion as to the benefits of rights really follow from what he has said? If it is assumed that rights are the sort of thing persons have whether or not any one can be said to be respecting those rights, the conclusion clearly does not follow from the fact that we are said to have rights. In fact, rights do not protect persons from arbitrary harms. No right protects me from being struck by lightning—nor does any moral right protect me from being electrocuted by a sadistic scientist. It is the scientist's behaving in that way which we may call "respecting my right not to be electrocuted" that protects me. As others do elsewhere, Wasserstrom slips in the phrase, above, 'and in which rights are generally respected' as if it were of minimal importance. But, of course, it is precisely persons acting in this way that makes a society secure.

Realizing that rights talk is necessarily concerned with the actions of those we might call potential rights violators allows us to see that an important part of what we are discussing is the behavior of potential rights violators. It is misleading and can be injurious to insist that the existence of a right be established before an action is forbidden. To say that I ought not to punch you because it would hurt you is sufficient—there is no further need to invoke that alleged thing that is your right not to be punched or hurt. Clearly, much of what is to be taken into account is the harm caused by our actions and the significance we give to that harm.

This way of looking at rights talk makes it clear that a debate as to whether or not the members of a certain group—be it dolphins or ethnic minorities—have a given right is largely a debate over how those debating ought to behave toward members of that group. When we see rights in this way we understand that, more than the having of rights, it is what we call recognizing and respecting the rights of others which makes us persons, and raises us above "mere animals."

HEATHER J. GERT

Hamilton College

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

SUPERVENIENCE IS A TWO-WAY STREET

Despair over attempts to reduce the nonphysical to the physical has led some materialists to try a shortcut. The supervenience of all properties on physical properties is held by some to be all that a physicalist could want or need.[1] I am suspicious of the shortcut. I do not think that the supervenience of the moral, the mental, the aesthetic, etc., establishes the primacy of the physical. I shall attempt to show this by demonstrating that the physical is not unique with respect to supervenience. Although it is true that the nonphysical supervenes on the physical, I argue that it is equally true that the physical supervenes on the moral, the mental, and the aesthetic. 'No difference without a physical difference' is an excellent slogan. The gist of this paper can also be summarized with slogans. 'No difference without a moral difference', 'no difference without a mental difference', and 'no difference without an aesthetic difference', are as (trivially) true as the physicalist slogan. Pluralists, like Nelson Goodman, could interpret this result as an argument for the multiplicity of worlds, but I take it merely to show the insignificance of supervenience.

Not every physicalist uses the alleged universal supervenience on the physical to justify physicalism. I have no qualms about those who use the concept merely to clarify the doctrine of physicalism. In fact, I think that Jaegwon Kim's[2] concept of strong supervenience has something to contribute to our understanding of the issue. My objection is to those who use supervenience as a premise.

The concept of supervenience entered the philosophical world via a discussion of moral properties.[3] G. E. Moore held moral properties to be both real and irreducible. In order to make them a bit less

[1] Among those who argue in this way are W. O. Quine, "Goodman's Ways of Worldmaking," in *Theories and Things* (Cambridge: Harvard, 1981); John F. Post, *The Faces of Existence: An Essay in Nonreductive Metaphysics* (Ithaca: Cornell, 1987); and John Haugeland, "Ontological Supervenience," *Southern Journal of Philosophy*, XXII, Supplement (1984): 1–12. Current work on supervenience owes much to the seminal papers of Geoffrey Paul Hellman and Frank Wilson Thompson, "Physicalism: Ontology, Determinism, and Reduction," this *Journal*, LXXII, 17 (October 2, 1975): 551–564, and "Physicalist Materialism," *Nous*, XI, 4, (November 1977): 309–345.

[2] "Supervenience and Supervenient Causation," *Southern Journal of Philosophy*, XXII, Supplement (1984): 46–50.

[3] Kim, *op. cit.*, traces supervenience back to Moore's "The Conception of Intrinsic Value," in *Philosophical Studies* (London: Routledge & Kegan Paul, 1922).

0022-362X/90/8712/695–701 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

"queer," he added that moral properties supervened upon and hence did not challenge the ontological primacy of the physical.

Given this historical background it is appropriate then that I stand this argument on its head. I shall endeavor to show that the physical supervenes on the moral just as much as the moral supervenes on the physical. If, pace Goodman, we assume that the physical and the moral cannot be equally primary, this demonstrates the metaphysical insignificance of the supervenience relation.

For a class of properties *S* (supervenient) to supervene on a class of properties *B* (base) is for it to be true that for any two things to differ in their *S* properties is for them to differ in their *B* properties.[4] But the distinction intended between *S* properties and *B* properties requires that there be differences in *B* properties which are not paralleled by *S* differences. Otherwise, not only does *S* supervene on *B* but *B* will also supervene on *S*. Some of the differences in *B* properties of pairs in the domain must have *S* differences that parallel them, namely, those which parallel the *S* differences themselves. What is needed is for there to be enough *B* differences to account for all the *S* differences but not enough *S* differences to account for all the *B* differences.

There are certain consequences of this definition of supervenience which I think have gone unnoticed. If every pair in the domain has a *B* difference, then the supervenience of *S* on *B* follows immediately. And if every pair also had a *S* difference, then *B* would also supervene on *S*. In a possible world in which every object has a unique color, then every property supervenes on color. This follows because in such a world there are no differences in shape, size, texture, etc., without a difference in color.

The apparent asymmetry of the supervenience relation, I contend, is an artifact of certain contingent features of our actual physical and moral languages. Actual physical language is far subtler and closer to an ideal physical language than actual moral language is to an ideal moral language. Actual physical language has incorporated mathematics, while actual moral language, despite the efforts of Jeremy Bentham, has not. This results in a vast preponderance in the number of physical distinctions we can make in the actual physical language compared to the actual moral language. Mathematics is a tool

[4] It is impossible to give a definition of supervenience that will satisfy all participants in the discussion. I shall try to prove the symmetry of supervenience with this definition. Others will have to judge whether or not my arguments hold for each of the current definitions in the literature.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

that enables the user to make literally infinite distinctions in as many dimensions as can be mathematicized. Our physical language permits infinitely fine gradations in mass, velocity, location, and many other variables.

Physicalists must be willing to consider relational physical differences in the base properties.[5] Compare Joe on earth with Joanne on twin earth. Joe believes that water is wet; Joanne believes that twin-earth water is wet. Yet there is no internal physical difference between the two. Both subvocalize 'water is wet' at appropriate times, but Joe speaks English and Joanne speaks twin-earth English. The difference in their respective beliefs supervenes on and is explained by the difference in their immediate physical environments, full of H_2O and XYZ, respectively, not by any difference in their internal physical states.

The base properties must include relations to noncontemporaneous events. Being United States currency does not supervene on the quality of the paper or the pattern of ink marks on the paper. It supervenes on who printed it. A perfect counterfeit would not be valid United States currency. Being money supervenes on relational physical properties connected with the causal history of the tokens in question. So not only do base properties include relations, they do not even have to be contemporaneous with the supervenient property.

Moreover, physical properties are often, if not always, dispositional. Inertial mass, for example, refers to a body's disposition to resist changes in its state of motion. Two bodies could differ in their inertial masses even if neither ever suffered any changes in their states of motion so long as counterfactually they would differentially accelerate if subjected to equivalent forces.

Finally, the base properties need not be detectable. "The physical base may be very complex and may not even be humanly discoverable; as a result it may be *unavailable* for a physicalist reduction or explanation of the psychological state. But it *must* exist if the mental strongly supervenes on the physical."[6]

So the base properties upon which moral properties supervene include relations, noncontemporaneous properties, and disposi-

[5] Paul Teller, "A Poor Man's Guide to Supervenience and Determination," *Southern Journal of Philosophy*, XXII, Supplement (1984), p. 139, explains clearly why relations must be included in the base properties.

[6] Kim, p. 49. As I have mentioned above, Kim is elucidating the supervenience relation, not using it to argue for the truth of physicalism.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

tional properties, none of which need be detectable. Moreover, these properties are mathematicized to allow an infinity of distinctions to be drawn. Recall that, if every pair of things differs physically, then the supervenience of everything on the physical trivially follows. The capacity for drawing distinctions inherent in this generously endowed physical language all but guarantees that this will, in fact, be the case.

The idealization of physical language, and especially its mathematicization, is practically sufficient to establish the supervenience of all other properties on the physical. It is quite proper for philosophers to consider idealizations of the physical language in metaphysical contests. I do not fault this practice, I applaud it. But idealization trivializes the supervenience relation, because it allows infinite distinctions to be made.

If there is a physical difference between every pair of individuals, then trivially there is no difference without a physical difference. But the mathematization of physics practically ensures that there is a physical difference between every pair of individuals. This can be demonstrated even in very simple possible worlds. Consider a world consisting of two peas in a pod. With only a crude qualitative physical language at our disposal, we should be forced to say that they were identical. Both are round, small, and green. But if we mathematicize, then we practically guarantee that there will be a physical difference. If none is detectable by human sense organs, we will use instruments. If none is detectable by instruments, we will perfect or hypothesize better instruments. One thing is certain: we will never be forced to admit there is no physical difference. When we consider the complexity of the actual world, plus the fact that the supervenience relation is supported by undetectable differences, relational properties, and dispositional properties, the mathematization of physics ensures supervenience. To repeat my central claim, it is the expressive power of physical language which allows us to assume physical differences where none are detectable.

If we idealize moral language in this way, then everything will supervene on the moral. If we consider undetectable dispositions, relations, and noncontemporaneous moral properties and especially if we make quantitative moral distinctions and not merely qualitative ones, we ensure that there will never be a physical difference without a moral difference.

I do not think anyone can honestly object to idealizing moral language in this context. In casuisty, we ought to use our actual moral language such as it is. But in metaphysics, we are trying to look

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

at the universe from a "God's eye" point of view and idealization is appropriate. Just as it is unfair to allow one team to use a different set of rules from another team, it is inappropriate to compare actual moral language to idealized physical language. The illicit comparison of an anemic qualitative moral vocabulary with a robust quantitative physical vocabulary creates the illusion of a metaphysical primacy for the physical. Let both teams play by the same rules and the result is a draw.

We should note the import of these results for the supervenience of moral on physical properties. The intended distinction between supervenient and base properties requires that no two things be morally different and physically identical, but that some things must be morally identical and physically different. Otherwise, moral differences would not only supervene on physical differences but physical differences would also supervene on moral differences, which would destroy the intended ontological primacy of the physical. But if there is always some moral distinction to be made, however small, between any two elements in the domain, the distinction between supervenient and base properties disappears. It is my contention that we can make moral distinctions as readily as physical distinctions, and so the supervenience argument for physicalism collapses.

Granting the same advantages to moral language produces parallel results. Consider an idealized consequentialist moral language in which the goodness of an act is equivalent to the degree to which it tends to promote human flourishing. Let 'tendencies' and 'human flourishing' be mathematicized. Do not restrict moral distinctions to human powers of measuring or detecting degrees of human flourishing. Allow the relevance of relational predicates to individuals distant in time and space to the acts in question. Finally, let the differences range over dispositions to affect flourishing and not just actual flourishing. Is it not apparent that we can always find some moral difference between any two acts?

It will not suffice to find two things that differ physically and share some moral quality, e.g., permissibility. This would indeed be easy to do. Tying my shoes and eating my breakfast are both permissible and quite distinct physically. For any two actions to be morally identical, however, it is not sufficient for them to share one moral property, they must share all moral properties. Although tying my shoes and eating my breakfast are both permissible, they are not necessarily morally indistinguishable.

There seem to be obvious counterexamples to my claim that there are no physically different and morally identical actions. Surely, it

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

will be argued, a trivial physical difference like the color of the victim's eyes can make no moral difference. It would be monstrous to hold that the murder of a blue-eyed victim differs morally from the murder of a brown-eyed victim. The sanctity of human life and the essential equality of persons requires that we hold all murders to be equally wrong.

My response to putative counterexamples of this sort is that, though I must agree that the color of the victim's eyes cannot make a difference with respect to whether or not the murder is wrong, it can nevertheless be accompanied by a moral difference. All murders are wrong. This is a simple moral truth which should not be denied. Do we need to admit that all murders are morally equivalent, however? I do not think that we must or that, in fact, we do admit this. The differences in the punishments we impose for different murders show that we believe murders to be not only morally different but even morally distinguishable. The sanctity of human life notwithstanding, some murders are more heinous than others.

Is it implausible to assume that Susan's killing blue-eyed Bill has some slight degree of difference in its tendency to promote human flourishing from Arnold's killing brown-eyed Betty? It is our habit of restricting moral distinctions to qualitative, detectable differences which leads us to the belief that all killing is morally the same. A quantitative moral language would lead to a different result. Surely all actions differ in some degree in their tendency to affect human flourishing.

At this point, we are ready to handle the toughest alleged counterexamples. Take the smallest possible physical difference and make it as remote as possible from human flourishing. For example, add one hydrogen molecule to some distant uninhabited star. What moral difference does this hydrogen molecule make? Clearly, not much. But it does not have to make a significant or detectable or even an actual difference; it only has to have some slight disposition to affect human flourishing. The extent of human knowledge is part of human flourishing, and by adding to the mass of a distant star it makes that star slightly more detectable by astronomers, promoting to this insignificant degree to the likelihood of human flourishing. This is so, even if no one ever actually detects the star. My example is admittedly extreme, but the point is to show that, if the moral is given the same latitude as the physical, then even the most unlikely counterexamples can be handled.

A similar case can be made for the supervenience of the physical on suitably strengthened aesthetic or psychological languages. Give

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

these languages the same resources for generating distinctions as the physical now has, and both will constitute a set of base properties upon which everything may supervene. Although we are certainly not used to doing so, we could employ a mathematicized aesthetic language and in this language express all the incredibly fine aesthetic differences that parallel physical properties. Even ordinary mortals are sensitive to the beauty of the double-helix, but we are too crass to distinguish the variations in beauty of different strands of DNA. But God is not so insensitive. The subtle chemical differences in strands of DNA are used by us to explain various inherited characteristics. Could not a divine mind use subtle aesthetic differences to explain the same things?

Concerning the supervenience of the physical on the mental, we need only remember idealism. Berkeley identified what we would call the physical properties of tables, their shape, size, weight, etc., with how perceivers might perceive tables under different conditions. We might object to idealism, but surely every physical difference in a table corresponds to some dispositional difference in how some perceiver might perceive it. In fact, every physical difference in a table supervenes on a host of dispositional perceptual differences in how the table would look to observers in different positions. No matter how small we make the physical differences, we can always find matching phenomenal differences. If we make the physical differences fall below the threshold of human sense organs, there are always instruments. When all else fails, we can always postulate superhumanly sensitive perceivers to pick up the remaining slack.

Supervenience is a two-way street.

RICHARD B. MILLER

East Carolina University

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

BOOK REVIEWS

Beauty Restored. MARY MOTHERSILL. New York: Oxford University Press, 1984. vii, 438 p. Cloth $29.95.

This is a good book. It is written with wit and sensitivity. It takes the author a long time to formulate the leading thesis, but the three-hundred-forty pages leading up to the formulation are well-used. They serve to acquaint the reader with Mary Mothersill's deep, reflective thoughts about beauty, and they are a kind of introduction to Mothersill's philosophical sensibility. Mothersill's remarks are sometimes personal, sometimes almost confessional, and sometimes rambling. Some readers may find this an inefficient kind of philosophy, but I appreciate the animation it gives to the philosophical theses. The writing is not self-indulgent. Frequently it is simultaneously charming and stimulating. For instance,

> Again there are problems: can love *create* beauty, as it were, *ex nihilo*? Or should we say, in Socrates' phrase, that what love creates is 'apparent beauty'? I treat a toad with courtesy; the toad turns into a prince; I fall in love. But suppose nothing happens; could I come to love a toad? A prisoner might, or a child. Would a toad then be beautiful? Perhaps it is a matter of degree: Alcibiades came to like the way Socrates looked without ever, perhaps, being inclined to describe him as handsome (275).

The character of the book is virtually unique. In recent aesthetics (and ethics) there has been considerable interest in the philosophy of the eighteenth century, especially that of Hume and Kant. This interest customarily has found expression in commentaries and philosophical historical work. Mothersill seems to me to have fully absorbed the philosophical culture of those authors, combined it with the character of recent analytical philosophy, and then turned to the question of beauty. Her book contains chapters that are useful commentaries on Hume and Kant, and it contains valuable passing remarks on George Santayana, John Dewey, Plato, Aristotle, Nelson Goodman, and many others; but its controlling aim is to argue for a certain conception of beauty (and along the way to urge that this neglected eighteenth-century concern deserves a central place in the philosophy of art). The conception is developed within the constraints of two assertions which Mothersill calls the *first thesis* and the *second thesis*.

> The *First Thesis* (*FT*): There are no principles or interesting laws of taste

0022-362X/90/8712/702–8
© 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

A principle of taste would have the form
'Whatever has property ϕ is *pro tanto* beautiful . . .'
An interesting law of taste "would tell us what features of items in a specified class will be *pro tanto* a cause of pleasure to subjects suitably qualified . . ."
A principle of taste or an interesting law of taste "would provide deductive support . . . for the judgment of taste under its normative aspect" (85–92).
The *Second Thesis* (*ST*): Some judgments of taste are genuine.
"A judgment is 'genuine', then, if it is contingent, hence either true or false, if it is such as to admit testing by anyone who cares to take the trouble and if there are determinate confirmation procedures that can be sketched in advance" (164).

Much of the book is Mothersill's supple discussion of the relation between FT and ST, and of what that relation has been taken to be. FT and ST readily appear to be inconsistent with one another. Thus Hume, determined to maintain a version of ST, is tempted to deny FT. Kant, on the other hand, absolutely committed to FT, essentially denies ST. Mothersill's own opinion is that both FT and ST are true—so obviously true, in fact, that neither needs an argument—and she takes on the burden of showing that ST is highly plausible despite the fact that FT makes it impossible to vindicate a judgment of taste by "deducing" it. I think she succeeds well enough, but I cannot do justice to her subtle argument without reproducing the sixth chapter, and so I shall just recommend those thirty-one pages to you. I shall note that one of her central insights, one she shares with Hume and Kant—particularly Hume—is especially compelling. It is that there is an essential connection between one's feeling of pleasure and one's judgment that something is beautiful, but not all feelings of pleasure are signs of beauty, and that the discovery that there is beauty in the world is made by discovering something about one's own feelings of pleasure. This is to say, or at least to suggest, that the reflective sense of the special pleasure of beauty arises, not from acutely epistemological worries about "verification" and "interpersonal judgments," but from analytical introspection. This introspection leads to close consideration of the object, and to the thought of others who may encounter the object, but it is within the introspection that the hint of beauty appears. I think this is a very good idea, and I am especially taken with Mothersill's ways of putting it.

> The work of criticism begins with the attempt to make it clear what it *is* in a particular item that, as it seems to me, accounts for the fact that it pleases me (166).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

On my account, the judgment of taste is a response to what I have called a 'notable' pleasure. A notable pleasure excites reflection and thought. . . . One begins to focus on the cause of one's pleasure and perhaps on the particular aspects of the item that make it exceptional. To take something to be beautiful requires, on my account, identification of an individual and the formation of a hypothesis to the effect that it is in virtue of its aesthetic properties that the individual pleases. A judgment of taste, even if never explicitly formulated, is somewhat self-conscious and formal (375–6).

In summary: on my view, the difference between finding something agreeable and taking something to be beautiful is not correlated with differences between types of items enjoyed, say food and wine as opposed to scenery and music. Nor does it correspond with the distinction between the pleasure of gratified desire and 'free' pleasure. It signalizes a difference, one might say, of attitude on the part of the subject. If his enjoyment of, say, background music in a cocktail lounge, is entirely casual, he will not bother about correctly identifying the cause of his pleasure nor about determining which particular aspects of the case affect him. . . . By contrast, someone who takes an item to be *beautiful* has been struck by something that captures his attention and makes him think; he believes, truly or falsely, that the item has extraordinary qualities and that it will be a cause of pleasure to anyone who attends to them (376–7).

Mothersill's rationalized, formalized account of beauty is carried in three definitions. These definitions, in turn, require definitions of 'aesthetic property' and 'indistinguishability', which are credited to Sue Larson. Here they are:

(i) Two individuals, x and y, are indistinguishable if and only if a specified subject S is unable, under standard conditions of observation, to tell one from the other (343).
(ii) 'ϕ is an aesthetic property' means that for any x, if ϕ of x, then for any y, ϕ of y, if and only if y is indistinguishable from x' (344).
Definition 1: Someone takes an individual to be beautiful if and only if the individual pleases him and he believes that it pleases him in virtue of its aesthetic properties (342).
Definition 2: Someone finds an individual beautiful if and only if the individual pleases him and he believes that it pleases him in virtue of its aesthetic properties and his belief is true (347).
Definition 3: Any individual is beautiful if and only if it is such as to be a cause of pleasure in virtue of its aesthetic properties (347).

I am unpersuaded by the analysis afforded by these definitions, but it is difficult to settle questions about the salience of this analysis, and it is critical to understand the point of the analysis. Mothersill says that "the aim of analysis is not to *recreate* (describe what it is

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

like) to find something beautiful, but to make articulate the commitments that our claims involve" (345). Before I note what seem to me some misrepresentations entailed by the analysis, let me note a powerful and subtle rationale underlying the analysis.

A principal idea lying behind Mothersill's analysis is one given by Arnold Isenberg in his "Critical Communication." (I could not be more in agreement with Mothersill's assessment of the great value in Isenberg's essay, now more than forty-years old and still not well-appreciated.) Isenberg noted that, when one refers to some property of an object by way of giving a reason for one's liking the object, one could not be alluding to a sound argument that took as its major premise the assertion that one likes all objects possessed of that property, because that assertion inevitably will be false (this is a version of FT). What one must be doing, then, according to Isenberg, is identifying the property-*as-it-appears-in-this-object* as the ground or the cause of one's pleasure. (Isenberg's leading example is the property of being a steeply rising and falling curve as it appears in a painting by El Greco, *The Burial of Count Orgaz*.) Mothersill has refined and formalized this idea: it is meant to be captured by her definitions. As Mothersill says, in explication, "The proposal of *Definition 1*, then, is that in taking an item to be beautiful I ascribe the pleasure that it gives me to its aesthetic properties, that is, to those properties which define the set of items such that I, at least, cannot distinguish any of them from the original item" (345).

If I am to understand the object to be the cause of my pleasure, on account of which I take the object to be beautiful, then there must be some causal law (regularity) linking properties of the object to my feeling. Mothersill's definitions provide for this, but at the same time they block the unwanted inference to the false assertion that some other object must also count as beautiful because of its possession of the same property: there is no *other* object, at least so far as I can tell: I believe that any other object in possession of the property in question will seem to me not to be an other object. And if I am right in taking the object to be beautiful, that is, if Definition 2 applies as well as Definition 1, then in fact I shall never notice the property except in this object.

I count my judgment of the object a genuine judgment (thereby exemplifying ST) because it does admit of testing, after a fashion. What I must do to persuade you of the truth of my judgment of beauty is get you to discern the (aesthetic) property in question, and trust that, once you do, you too will feel pleasure in virtue of that property. There is no *proof:* if I do not secure this "community of

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

feeling," I shall not be in a position to indict you for ignorance, blindness, stupidity, or irrationality, but this is just to hold to FT.

Thus, Mothersill's definitions constitute a strong, penetrating analysis. I cannot yet accept it because I am troubled by what seems to me its overstrong requirement, the implied requirement of indistinguishability. Here are two partial examples meant to show that the requirement is too strong. I do not think they simply make the case against Mothersill's analysis: the reader will have to see how they strike him.

1. Suppose the El Greco painting, *The Burial of Count Orgaz,* is beautiful. Then, according to Definition 3, this painting is such as to be a cause of pleasure in virtue of its aesthetic properties. Suppose you are someone who finds the painting beautiful, and thus that the painting does please you in virtue of its properties. According to Definition 2, then, you correctly believe that these properties are aesthetic properties. Suppose that there is another picture, a copy of *The Burial of Count Orgaz,* which is also beautiful and is also found beautiful by you. Call the original painting P and the copy $P(C)$. Suppose that the relevant property in both cases is one that could be referred to by 'steeply rising and falling curve'. One of these properties is a steeply-rising-and-falling-curve-as-it-appears-in-P, and the other is a steeply-rising-and-falling-curve-as-it-appears-in-$P(C)$. The question is: Are these properties the same? If so, then the property is not an aesthetic property *if you can distinguish P from* $P(C)$. And if the property is not an aesthetic property, then you do not, after all, find either P or $P(C)$ beautiful. It seems to me entirely possible that the same beauty-making property appear in more than one individual and that these individuals be distinguishable by you. The Mothersill theory cannot permit this, because if the same property appears to you in more than one distinguishable individual, then that property, by definition, is not an aesthetic property, and you do not find an individual to be beautiful unless you correctly believe it to be pleasing you in virtue of some of its aesthetic properties.

This is not an objection to Mothersill's conception of an aesthetic property. Such an objection has been made by some commentators on Mothersill's theory, and I once made it myself, but it is a pointless objection. Mothersill has defined 'aesthetic property' in this way, and there is no objection to that. (In fact, her conception seems to me an advance over Frank Sibley's more familiar conception, because the Mothersill-Larson conception gives genuine content to the conception.) The objection is that her conception of beauty, depending as it does upon this conception of aesthetic properties, is not faithful to

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

the commitments that our claims about beauty involve. I am saying that you might well find yourself wishing to claim that two (distinguishable) individuals were both beautiful, and both in virtue of their common possession of some property. This is not to say that you are now committed to what Isenberg says is false, a claim that, whenever a steeply rising and falling curve appears in a painting, you will be pleased by that painting. Isenberg may well be right, that there is no sheerly general description that will articulate the relevant property in such a way that it can be appreciated outside the context of the individual in which it appears; but it may also be true that this single highly specific property appear in more than one individual. Why not?

One might reply that the property as it appears in P *must* be different from the property as it appears in $P(C)$. One would certainly make this reply if one believed that every individual embodied exactly one big aesthetic property. [Mothersill does not believe this. In fact, she says that an individual may well possess more than one aesthetic property (354).] I cannot see a non-question-begging reason for making this reply. Then what reply can be made? If there is none, we are left asking how to decide when a pleasure-causing property of an individual is a ground for our taking the individual to be beautiful. If the property does not have to be an aesthetic property, as defined by Mothersill, then what does it have to be? It will have to be an "Isenberg property," but it remains to say just what such properties are.

2. Suppose that both you and I seem to find P beautiful, and that we both also seem to find $P(C)$ beautiful, and that we both are pleased by both individuals in virtue of their respective rising and falling curves. But suppose that you can distinguish P from $P(C)$ and I cannot. [To avoid an irrelevant possible complication, suppose that the way you distinguish P from $P(C)$ has nothing to do with the relevant curves.] We are now in the curious position of having to say that I am the only one who finds both individuals beautiful, although it now transpires that in fact I think there is only one individual. It seemed that you did as well, and that you did so in about the same way as I did, but now we see that, in fact, the relevant property is an aesthetic property for me but not for you. Thus, Definition 2 compels us to say that you did not find either individual beautiful. It seems to me more plausible to say that we both find, say, P beautiful in exactly the same way.

I do not suppose either of these examples to be decisive. They depend upon how plausible they seem upon reflection. If they are

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

plausible, then Mothersill's conception of beauty needs an adjustment. Even if that is true, the book remains valuable from start to finish, and I am pleased to recommend it.

TED COHEN

University of Chicago

Contingency, Irony, and Solidarity. RICHARD RORTY. New York: Cambridge University Press, 1989. xvi, 201 p. Cloth $34.50, paper $10.95.

Each of the words in the title names a major theme of this book. It begins with accounts of what Richard Rorty calls vocabularies, of truth and of language in general. Examples of alternative vocabularies—Rorty also calls them language-games—are "the vocabulary of ancient Athenian politics versus Jefferson's, the moral vocabulary of Saint Paul versus Freud's, the jargon of Newton versus that of Aristotle, the idiom of Blake versus that of Dryden." Of these, Rorty says that they are not made true or false by "the world" (5). Truth is not a relationship of representation between a set of descriptions in such a vocabulary and something else—reality, the world, or whatever. "To say that Freud's vocabulary gets at the truth about human nature, or Newton's at the truth about the heavens" is to pay "an empty compliment" of the kind "traditionally paid to writers whose novel jargon we have found useful" (8).

Vocabularies are useful sets of metaphors adopted or discarded insofar as they serve purposes that engage "us." Such changes do not involve criteria or choice. "Europe did not *decide* to accept the idiom of Romantic poetry, or of socialist politics, or of Galilean mechanics. . . . Rather, Europe gradually lost the habits of using certain words and gradually acquired the habit of using others" (7).

This understanding of vocabularies as useful sets of metaphors is presented as deriving from a view of language in general attributed to Ludwig Wittgenstein and Donald Davidson. What we are to learn from Rorty's Wittgenstein and Rorty's Davidson seems to be: that "there is no fixed task for language to perform" (13), that alternative vocabularies are like alternative tools with different uses (11–13), and that language is the contingent product of blind evolution.

What then is each of us to make of the particular language that he

0022-362X/90/8712/708–711 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

or she has inherited? We are to remake ourselves, using Harold Bloom's conception of "the strong poet" and Friedrich Nietzsche's of the will to self-overcoming, but these somewhat aristocratic conceptions are to be democratized by invoking Sigmund Freud, a Freud who showed that *everyone* has the same unconscious need as the strong poet, to remake his or her own self by redescribing in his or her own terms that tissue of contingencies that his or her past provided.

The final vocabulary of any person is that in which long-term projects are formulated, deep hopes and fears are expressed, and the story of one's life is told. It is final in that its use cannot be defended against rivals by noncircular arguments. The ideal citizen of a liberal society—and Rorty presents the vocabulary of self-creation as what liberal societies now need—will be what Rorty calls an ironist in respect of his or her final vocabulary. Rorty's ironist uses one particular final vocabulary, but is aware of alternative vocabularies and experiences consequent doubts about his or her own, recognizing both that no rational argument can settle the issues between rival vocabularies or allay his or her doubts, and that his or her own vocabulary is not "closer to reality than others" (73). The ironist takes as moral adviser a type of literary critic who has learned that the importance of G. W. F. Hegel, Søren Kierkegaard, and Nietzsche is that they helped to "de-cognitivize, de-metaphysicize philosophy" (79) and so enabled us to read philosophical texts as literary texts, reading in illuminating juxtaposition authors hitherto treated as literary with others hitherto treated as philosophical, so that we revise our vocabulary and in so doing revise our own moral identity. "Literary criticism does for ironists what the search for universal moral principles does for metaphysicians" (80).

One central task is to foster not only appreciation of the tasks of self-creation in private life, but also what Rorty takes to be a peculiarly liberal abhorrence of cruelty in public life, particularly by sensitizing to hitherto unnoticed types of cruelty and humiliation. Our obligations are not founded upon rationally defensible universal principles; they express our social solidarities and arise from forms of practice in which we participate. Where in the philosophy of language Rorty had invoked Wittgenstein and Davidson, in moral philosophy he invokes Michael Oakeshott and Wilfrid Sellars. And where, in extending his account of the tasks of self-creation through innovations of vocabulary, he appealed to Marcel Proust, Nietzsche, Martin Heidegger, and Jacques Derrida, here he appeals to Vladimir Nabokov and George Orwell.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Thus summarized, Rorty's impressive themes plainly need long and detailed spelling out. They do not receive it in this book, written as it is in intellectual shorthand whose compressions, allusions, and flag-waving, signposting rhetorical style make it too easy a victim for any initially unsympathetic reader. Far too much work is left to the reader: there are argumentative gaps to be filled, places where everything turns on the detail, but detail is absent, incoherences to be resolved. Three questions stand out.

The first concerns the true place of justificatory philosophical argument in Rorty's constructions. On the one hand, he abjures it: philosophers should not be asked for arguments against the views that Rorty rejects, for such arguments are bound to be question-begging (8–9); "I am not going to offer arguments against the vocabulary I want to replace" (9); "the ironist thinks that such arguments are . . . useful as expository devices, but in the end not much more than ways of getting people to change their practices without admitting that they have done so" (78). On the other hand, at key points Rorty uses what he elsewhere abjures. He quotes and extends an argument of Davidson about what *must* be involved in any substantial change in our views or values (49); he treats Freud's views as well-founded (31–2); and he argues sporadically against the thesis that it is possible to identify conditions of possibility for types of judgment or experience (e.g., 125).

Secondly, it is not clear how we should understand Rorty's account of the ironist's final vocabulary. What is it about such vocabularies that deprives them of the possibility of noncircular justification and of rational defeat at the hands of their rivals? On this Rorty is silent. Yet rival incommensurable schemes of thought and practice have sometimes developed in ways that made it rational by the standards of either to discard one and adopt the other. And there are well-known arguments supporting the conclusion that beliefs that cannot be defeated somehow or other fail the test of rationality. Rorty doubtless takes these considerations to be irrelevant. But we need to learn why.

Finally, Rorty's ambivalence about philosophical argument renders quite unclear the point of his appeals to Wittgenstein, Davidson, and others. He cannot, it seems, be offering us grounds drawn from their theories; but if he is offering us their conclusions detached from any rational grounds, why should we be interested? This difficulty is aggravated by the way he rewrites his culture heroes in the course of invoking them: Rorty's Wittgenstein means by 'language-games' something notably different from what Wittgenstein

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

meant; Rorty's Freud is a perspectivalist, unlike Freud himself; and Rorty admits to having reinvented Hegel. This multiplication of fictions suggests that Rorty's present attitudes will find adequate expression only in some genre more remote from those of both traditional and contemporary philosophy than he is here prepared to recognize. Inside these didactic expositions there is perhaps a novel pleading to be let out.

ALASDAIR MACINTYRE

University of Notre Dame

Reconceptions in Philosophy and Other Arts and Sciences. NELSON GOODMAN and CATHERINE Z. ELGIN. Indianapolis: Hackett Publishing Company, 1988. xiv, 174 p. Cloth $21.50.*

This welcome book consists of ten chapters, five by Nelson Goodman, three by Catherine Z. Elgin, and the rest, jointly authored. Both in substance and in elegance of style, the work is clearly the result of extensive collaboration. Although the chapters can stand on their own—and indeed, each of them, save one, is eventually to be published elsewhere—together they display a rich thematic and argumentative unity.

The seven chapters in "Part Two: Exploration" apply and advance Goodman's theory of symbolism as first offered in *Languages of Art.*[1] They range over a wide variety of topics, including meaning and reference in architecture, the identity of works, the concept of variation in the arts, imagery in psychology, comparisons between linguistic and pictorial competence, the conditions for representation, and distinctions between facts and conventions, analog and digital systems, and the pictorial and verbal.

Where there is a notational system permitting works of art to be identified syntactically or semantically, the works are "*allographic.*" In particular, texts are allographic, so that, in chapter III, Goodman and Elgin argue that many interpretations of a text are multiple interpretations of the *same* work. For the text can be precisely identified independently of an interpretation or version.

* My thanks to Catherine Elgin and Nelson Goodman for their comments, which saved me from a number of errors and misreadings.

[1] Indianapolis: Hackett, 1968, 2nd ed.

0022-362X/90/8712/711-6 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Musical notation being, like a text, allographic, compliance with a score determines the identity of the work. So a variation of a theme in music must offer a different score. (That is not sufficient, however; see 69–73.) The central question Goodman raises is: "What special relationships of likeness and difference must obtain between two passages for one of them to be a variation of the other?" (67). His answer involves intricate explorations that begin with the proposal that functioning as a variation depends upon the way one passage *refers* to another. The account of variation is briefly extended to other arts (e.g., printmaking).

The climax of chapter IV ("Variations on Variations—or Picasso back to Bach") is Goodman's discussion and comments upon Picasso's famous variations on Velazquez's *Las Meninas,* 22 of which are reproduced (as is the Velazquez). Although there is much to be learned from all the chapters in part two, this one marks a highpoint. Goodman guides us with his theory of symbols (and expert's eye) on a tour of the art discussed and reproduced, the reader coming away exhilirated, his cognitive and emotive experience enriched.

Broad Goodmanian views surface in many of the chapters, but most of the specific analyses can be appreciated without endorsement of nominalism, irrealism, rigorous relativism, antifoundationalism (and coherentism), rejection of natural kinds and the given, or world pluralism. Other Goodmanian ideas, aside from the theory of symbols, play a more substantial role: the reflective equilibrium model of justification, the rejection of resemblance as a criterion for representation, secondary extensions, projectibility, and entrenchment. The optimistic and open-ended spirit of system building also features prominently: "Unlike terminal skepticism and irresponsible relativism, constructionalism always has plenty to do" (166).

Evident throughout is the Goodmanian delight in rejecting the standard oppositions and alternatives. Thus, in part one, which provides a clear overview of Goodman and Elgin's research leading up to this volume, Elgin begins:

> The theory of knowledge to be sketched here rejects both absolutism and nihilism, both unique truth and the indistinguishability of truth from falsity. It stresses reconstruction over deconstruction, and tolerates neither the noumenal nor the merely possible nor any ready-made world (3).

Again, in a Rylean spirit of ontological deflation and clarification, Goodman appeals in chapter V ("Sights Unseen") to his notion of secondary extensions to generate a fictionalist view of mental imagery.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Goodman writes: "we are stuck with both horns of the dilemma: that we can talk significantly about mental images and that there are none" (85). And in the next chapter, "Inertia and Invention" Goodman concludes:

> In sum, I have been urging such obvious points as that there is no firm distinction between fact and convention but that that distinction is very important; that the line between fact and convention shifts often and may be drawn anywhere but is not capricious; that when a convention (as option) becomes a convention (as the usual) it thus tends to become factual; and that rather than the facts determining how we take them, how we take them determines the facts—but that we had better be careful how we take them (100).

The first of the two chapters ("The Epistemic Efficacy of Stupidity") in "Part III: Premonitions" is an indirect way of arguing against the agenda of mainstream epistemology, thus serving as an introduction to the "reconception" proposed in the closing chapter. It is "indirect" in that it sets up a problem for various analyses of knowledge, and then, rather than provide her own solution, Elgin claims that the problem itself is symptomatic of what is wrong with the enterprise.

The problem is, roughly, that greater knowledge of a subject itself can undermine a person's knowing more statements within that subject. That can occur through the person's offering finer judgments, or the greater knowledge either rendering relevant more potential counterevidence or undermining the person's confidence. So Elgin wants to conclude that, for these analyses, "The quest for knowledge may then be furthered by the cultivation of obtuseness" (136). So,

> . . . knowledge, as contemporary theories conceive it, is not and ought not be our overriding cognitive objective. For to treat it as such is to devalue cognitive excellences such as conceptual and perceptual sensitivity, logical acumen, breadth and depth of understanding and the capacity to distinguish important from trivial truth (152).

The thesis is challenging, but there are two difficulties. The first is with details of her criticisms. For example, a person with refined knowledge of wine might in a taste test "confuse a Margaux with a St. Julien," (141) and so would not know that the wine he is tasting is a Margaux (even if it is and he believes it is and . . .). Whereas someone else, who lacks such refinement, could make a correct judgment about this wine which did constitute knowledge (e.g., 'It's vintage stuff'). I do not find anything particularly problematic about this

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

result. Trivially, finer judgments are bolder and bolder ones, riskier. [A point that Elgin seems to accept stated in general (151–2), though she still takes this particular consequence of different analyses as objectionable.] Moreover, we cannot conclude that the expert knows less than the man of unrefined taste simply from this difference alone. Among other statements that it is reasonable to imagine that the expert knows but his less refined friend does not is: 'Either this is a Margaux or a St. Julien'.

But the main difficulty for Elgin's argument is her assumption that these analyses of knowledge are committed to a rather direct link between the truth conditions for knowledge and cognitive advice. Thus, in the discussion of the above case, she concludes: "So as we refine our categories, we diminish our prospects for knowledge" (141). And although she does not say so explicitly here, her ultimate conclusion (quoted above from 136, 152) indicates that she wants us to infer that, on these analyses, "We therefore shouldn't refine our categories, in order not to diminish our prospects for knowledge." But I do not find cognitive advice of this kind coming from those offering these analyses, and I do not see why they are committed to maintaining anything like the strong claim that knowledge ought to be "our overriding cognitive objective." Silence on the connection between what it is to know and the value of knowing permits only a temporary reprieve, however, and it is to Elgin's credit that she forcefully challenges this silence.

In chapter X, Goodman and Elgin offer "A Reconception of Philosophy." Aside from moving philosophers away from issues that they regard as dead ends, such as skepticism, their primary goal is to further the Goodmanian challenge to the separation of the arts and the sciences:

> We work *from* a perspective that takes in the arts, the sciences, philosophy, perception, and our everyday worlds, and *toward* better understanding of each through significant comparisons with the others (164).

The heart of their proposal is to replace or supplement three crucial epistemological concepts: certainty, knowledge, and truth. They propose in their place adoption, understanding, and rightness, respectively. "Adoption," Goodman and Elgin write, "is a matter of putting to work, of making or trying to make a fit" (159). As with entrenchment, the continued success of an adoption may lead to "an increasing claim to rightness." But adoption itself implies no degree of confidence. (Because of its failure to imply any degree of confi-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

dence, I am unsure how "current adoption" could replace, as they propose, Goodman's notion of "initial credibility." For the original purpose of the latter is to provide warranted, though neither certain nor permanent, stopping points to block the regress of justification.)

When we replace knowledge by understanding, we leave behind a concept that is a poor mix of flawed ingredients—namely, truth and certainty (or uncertainty)—whose pursuit leads us to one or another foundationalism. On their view, however, the cognitive endeavor,

> . . . taken as the advancement of *understanding* . . . . starts from what happens to be currently adopted and proceeds to integrate and organize, weed out and supplement, not in order to arrive at truth about something already made but in order to make something right—to construct something that works cognitively, that fits together and handles new cases, that may implement further inquiry and invention (163).

Rightness is a matter of symbols functioning well, and as such is not limited to statements. Even if we restrict our attention to statements, truth is too limited as an evaluative dimension, since a true statement that is irrelevant is wrong.

It is a provocative proposal, worth extended discussion. Consider, for example, how readily we take a critic's work as done when he shows that a premise of an argument is false. But has he really shown, what is no doubt suggested, that the premise is interestingly, irremediably, or importantly false? Yet, for Goodman and Elgin's reconception to get the attention deserved, it would help if they invoked it in standard ways as a basis for in-depth criticism and comparison with other philosophical positions and arguments. Only in chapter IX (and much less so, chapter VII on different competencies with symbols) are we able easily to locate their proposals in the midst of the ongoing discussions that most engage the profession. Goodman and Elgin are just more disposed to reorient through doing and modeling than criticizing and pronouncing. So we are left intrigued and puzzled, wanting more.

Goodman is one of the most highly regarded contemporary philosophers. For many years, he has pursued an audacious vision, a vision not only allowing him to see further, but one that illuminates in different directions, many unsuspected. In this volume, we benefit by gaining a clearer grasp of that vision through observing his theory of symbols hard at work, and through blunt, bold statements of how that vision should alter the dominant conceptions in and of philosophy.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

 Digitized by Arya Samaj Foundation Chennai and eGangotri

The Hackett Publishing Company continues to serve the philosophical community very well, producing works of high quality at reasonable prices.

JONATHAN ADLER

Brooklyn College, CUNY

Realizing Rawls. THOMAS W. POGGE. Ithaca: Cornell University Press, 1989. 296 p. Cloth $36.95, paper $12.95.

In the introduction to this important new book, Thomas W. Pogge characterizes the writings of John Rawls since the publication of *A Theory of Justice* as exhibiting a trend "toward abstraction, vagueness, and conservatism." He explains this tendency as being "due, in large part, to the widespread criticism of [Rawls's] work, which, in this decade at least, has been predominantly conservative" (4). Pogge's tone, throughout the entire work, is so respectful of Rawls, so sober and responsible, that I am not entirely sure he recognizes what a devastating condemnation of Rawls these observations contain.

The effect of Rawls's concessions, Pogge decides, has been to move "the debate in exactly the wrong direction" (4). Accordingly, he takes it as his central task to follow out the radical implications of Rawls's theory, and to exhibit their "progressive power" (9). His program, as I understand it, has three parts, corresponding quite closely to the three parts of his book: first, a response to Rawls's two most powerful and widely heeded critics, Robert Nozick and Michael Sandel, neither of whom, Pogge thinks, has laid a glove on Rawls; second, a close examination of the core of Rawls's theory in the light of all of his writings since *A Theory of Justice,* in an attempt to show that Rawls's shifts of doctrine and emphasis, all of which have been conservative, establishmentarian, celebratory retrenchments, are philosophically unwarranted and incompatible with the deeper intuitions of Rawls's own theory; and, finally, a true radicalization of the Rawlsian theory of justice by extending it, in ways that Rawls himself resists, to the global sphere, where, Pogge argues persuasively, application of the two principles of justice on *any* plausible interpretation would have throughgoingly revolutionary consequences.

In short, Pogge offers himself up as a Left Rawlsian, staking out a relationship to Rawls very much akin to the stance adopted by Marx to that other manic system builder, Hegel.

0022-362X/90/8712/716–720 © 1990 The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Pogge succeeds completely in this enterprise, in my judgment, partly, of course, because he is an acute, intelligent, knowledgeable philosopher (some years ago, I came into possession, privately, of elaborate study notes on Kant's *Rechtslehre* prepared by Pogge for people teaching Contemporary Civilization at Columbia—they were the best thing I have ever read on that subject), but partly also because the thesis he is defending is true—indeed, I would suggest, manifestly true. If one attempts to apply the difference principle seriously, for example, one immediately recognizes that it cries out for reductions in the gap between high- and low-wage jobs so dramatic and far-reaching as to constitute a social revolution.

One of the abiding oddities of Rawls's career has been the flagrant contradiction between the radical leveling implied by his doctrine itself and the apparent absence in Rawls, either as a philosopher or as a citizen, of any reforming impulse. If the advantaged position of those at the top of the income pyramid is *really* to be justified *solely* by its tendency to improve the lot of those at the bottom, then on any halfway plausible construal of the facts of economics, psychology, and sociology, some hair-raising redistribution, through either taxation or a national wages policy, is absolutely inevitable. Nozick was quite correct to respond to this challenge by claiming that the riches of the rich are theirs by an absolute, uninfringible right. Any defense less extreme would not stand a chance against the difference principle.

Pogge cloaks extremely powerful moral condemnations of Rawls's substantive judgments in language so hedged and cautious that the true power of his claims is liable to go unnoticed. Two examples will illustrate this point. First, while discussing Nozick's criticism of Rawls, Pogge writes, in a characteristically cautious fashion: "It is hardly obvious that the basic institutions we participate in are just or nearly just. In any case, a somewhat unobvious but massive threat to the moral quality of our lives is the danger that we will have lived as advantaged participants in unjust institutions, collaborating in their perpetuation and benefiting from their injustice" (36). He then adds, in a footnote: "This sentence, I realize, goes against everyone's favorite moral conviction, namely, that there is nothing seriously wrong, morally speaking, with the lives we lead." A good deal later on, in a section devoted to a devastating critique of Rawls's post-*A Theory of Justice* tendency to elevate considerations of formal political rights above substantive matters of economic equality, Pogge observes:

> It is not credible that in the United States today the lowest prospects for developing and exercising the two moral powers would be raised more

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

by overturning the "profoundly dismaying" [Rawls's words] precedent of *Buckley v. Valeo* (in which the Supreme court declared unconstitutional various congressional limits on election expenditures) than by improvements in the diet, shelter, or education of the poorest citizens (134).

Indeed! Pogge goes on to observe, again in a footnote, that "[r]emoving a restriction on religious freedom . . . will not improve the situation of those whose prospects for developing and exercising their two moral powers are blocked by severe poverty. Its removal would nevertheless, on Rawls' account, have priority over relieving their poverty."

I must confess that in the past I considered the lexical priority of the equal-liberty principle over the difference principle to be more a consequence of internal systemic needs and constraints than a genuine reflection of Rawls's moral sensibility. It seemed to me too harsh to impute to him, as principled reflective moral sentiments, such appalling views as those exampled by Pogge in the passages just quoted. But by a meticulous examination of the many essays, speeches, and other writings that Rawls has produced since the appearance of *A Theory of Justice,* Pogge demonstrates incontrovertibly, in my judgment, that Rawls really does believe what his elevation of the first principle over the second implies.

Pogge's strategy, in the first part of his book, is to protect *A Theory of Justice* by imputing to Rawls the most coherent, intelligent, and internally defensible version of it that he can, and then to show that the criticisms of Nozick and Sandel simply miss the mark. Pogge's discussion is far too detailed to be discussed in a brief review. Its core, as I understand him, is a distinction which *is* simple enough to be summarized, and about which he seems to me to be quite correct. Briefly, Pogge argues that the principles of justice have as their referent the institutional structure of society, not the pattern of particular outcomes resulting from that structure. In this way, we might say, Rawls's view is big league rather than farm league. In farm league (at least as I recall its local version from the days when my sons were little), even after absolutely fair rules had been adopted and enforced, if the outcome was wildly one-sided, the adults who were running things would adjust matters so as to soothe the feelings of the kids unlucky enough to be on the losing side. But in the big leagues, so long as the rules are fair, and are impartially applied, it is just tough luck if the Yankees dominate the American League, or the Reds the National League, for a decade at a time.

To be sure, adjustments in the rules may be adopted to compensate for a certain pattern of distribution of talents—the designated

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

hitter rule, for example, has the effect of adjusting for the fact that pitchers tend to be lousy hitters. And if—to change the example—an African basketball league were to be organized in which teams of Watutsi played against teams of Pygmies, it might be necessary to invent an adjustable backboard that could lower or raise the hoop depending on which team was driving for it. But once such systemic adjustments had been carried out, no player would have a claim of justice against the game on the grounds that he, for example, although a Pygmy, found it easier to sink the ball in an elevated than in a lowered hoop.

All of this is perfectly correct, and Pogge manages to defend Rawls quite successfully against Nozick and Sandel, thereby preserving Rawls, undamaged, for use against himself!

In part two, as I have indicated, Pogge shows by a detailed examination both of *A Theory of Justice* and subsequent writings that Rawls has systematically, progressively, and entirely implausibly moved from his original concentration on the difference principle to an almost complete emphasis on the equal-liberty principle (even elevating the equal opportunity half of the difference principle over the strictly redistributive half). The effect has been to drain the theory almost entirely of its critical power. In his most recent writings, Pogge shows, Rawls has become virtually a flack for the American way of life. Gone is that generous commitment to remedying the structural economic inequities visited upon the underclass, which attracted so many readers to the earliest versions of the theory. (By the by, Pogge adopts the familiar bibliographical convention of referring to Rawls's various writings by their initials. "Kantian Constructivism in Moral Theory" becomes KCMT. This produces a splendidly, but probably unintentionally funny footnote, in which Pogge quotes the following passage from Rawls, *a propos* the Pecksniffian concern that lazy types might choose not to work and then claim, on the basis of the Difference Principle, a redistributive share of the products of others' labor: "those who surf all day off Malibu must find a way to support themselves and would not be entitled to public funds." This Rawlsian version of the problem of welfare mothers driving Cadillacs is identified as having, as its source, PRIG 257n! which, it turns out, is a citation of a 1988 essay entitled "The Priority of Right and Ideas of the Good.")

The shortest, but clearly, Pogge thinks, most original part of the book is the third, in which his progressive, strongly redistributive reading of the two principles is extended to the global sphere. I shall not summarize Pogge's views here—they are not difficult to imagine—but I would like to say something about the inner logical tension in

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Rawls's theory between the parochial, social contractarian reading of it as applying only within the borders of a nation-state, and the internationalist reading that Pogge advocates.

Rawls's original idea, more than thirty years ago when he published "Justice as Fairness," was to construe the social contract on the model of a bargaining game, in the game-theoretic sense. It was thus formally built into the model of analysis that there were a multiplicity of parties engaged knowingly in interactions with one another. No sense at all could be given to the idea of a unanimous agreement on mutually advantageous principles without that supposition. In this way, Rawls successfully modeled the situation of a society of citizens deliberating, in an inward looking fashion, about the principles that would henceforward regulate their centripetal relations, so to speak. When he switched to the "veil of ignorance" version of his model, however, in response to certain internal logical failings of the original argument, Rawls effectively (and, I would judge, unwittingly) forfeited the notion of a bargaining game. Instead, he transformed his theory into a quasi-formal analysis of a problem in individual rational choice. Once this change had been made, Rawls no longer had any plausible ground for restricting his principles to the regulation of the internal institutional arrangements of an on-going society. In effect, the veil of ignorance accomplishes the sort of secular universalization that occurred when the Stoics extended Greek moral and political principles to the world stage of the Roman empire.

The merit in Pogge's discussion lies not merely in his exploration of what the Rawlsian principles of justice would look like on a world scale, but in his meticulous demonstration that Rawls himself really has no justification for resisting that transformation of his theory. *Realizing Rawls* establishes Thomas Pogge as an important new voice in the political philosophical dialogue. I strongly recommend it to the readers of this JOURNAL.

ROBERT PAUL WOLFF

University of Massachusetts/Amherst

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

NEW BOOKS

ALTMAN, ANDREW: *Critical Legal Studies: A Liberal Critique.* Princeton: University Press, 1990. ix, 206 p. Cloth $26.50.

ANDERSON, CHARLES W.: *Pragmatic Liberalism.* Chicago: University Press, 1990. xiii, 220 p. Cloth $27.50.

ANDRESEN, JULIE TETEL: *Linguistics in America, 1769–1924: A Critical History.* New York: Routledge, 1990. vi, 308 p. Cloth $36.00.

APPELBAUM, DAVID: *Voice.* Albany: SUNY, 1990. xiv, 155 p. Cloth $44.50, paper $12.95.

ÁRDAL, PÁLL S.: *Passion and Value in Hume's* Treatise. Edinburgh: University Press, 1989. xxxix, 220 p. Paper $13.50.

ARNOLD, N. SCOTT: *Marx's Radical Critique of Capitalist Society: A Reconstruction and a Critical Evaluation.* New York: Oxford, 1990. xviii, 334 p. Cloth $35.00.

ARONSON, RONALD: *"Stay out of Politics": A Philosopher Views South Africa.* Chicago: University Press, 1990. xiv, 164 p. Cloth $32.00, paper $14.95.

ATHERTON, MARGARET: *Berkeley's Revolution in Vision.* Ithaca: Cornell, 1990. xii, 249 p. Cloth $29.95.

AUSTIN, C. R.: *Human Embryos: The Debate on Assisted Reproduction.* New York: Oxford, 1989. x, 163 p. Paper $15.95.

BAKER, KEITH MICHAEL: *Inventing the French Revolution: Essays on French Political Culture in the Eighteenth Century.* New York: Cambridge, 1990. x, 372 p. Cloth $49.50, paper $16.95.

BALDWIN, THOMAS: *G. E. Moore.* New York: Routledge, 1990. xviii, 337 p. Cloth $93.00.

BANNER, MICHAEL C.: *The Justification of Science and the Rationality of Religious Belief.* New York: Oxford, 1990. x, 196 p. Cloth $49.95.

BARNES, JONATHAN: *The Toils of Scepticism.* New York: Cambridge, 1990. xiii, 161 p. Cloth $34.50.

BARRY, BRIAN: *Theories of Justice. Vol. I.* Berkeley: California UP, 1989. xv, 428 p. Cloth $45.00.

BENCIVENGA, ERMANNO: *The Discipline of Subjectivity: An Essay on Montaigne.* Princeton: University Press, 1990. xi, 132 p. Cloth $19.95.

BERTOLET, ROD: *What Is Said: A Theory of Indirect Speech Reports.* Boston: Kluwer, 1990. xv, 251 p. Cloth $79.00.

BISHOP, JOHN: *Natural Agency: An Essay on the Causal Theory of Action.* New York: Cambridge, 1990. xi, 211 p. Cloth $39.50.

BOURGEOIS, PATRICK L.: *The Religious within Experience and Existence.* Pittsburgh: Duquesne, 1990. ix, 132 p. Cloth $23.95, paper $14.95.

BOWIE, ANDREW: *Aesthetics and Subjectivity: From Kant to Nietzsche.* New York: Manchester UP, 1990. viii, 284 p. Cloth $59.95.

BRAITHWAITE, JOHN, and PHILIP PETTIT: *Not Just Deserts: A Republican Theory of Criminal Justice.* New York: Oxford, 1990. ix, 229 p. Cloth $48.00.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

BRANDON, ROBERT N.: *Adaptation and Environment.* Princeton: University Press, 1990. xi, 214 p. Cloth $29.95.

BROWN, CLIFFORD: *Leibniz and Strawson: A New Essay in Descriptive Metaphysics.* Munich: Philosophia, 1990. 120 p. Cloth $29.00.

BROWN, FRANK BURCH: *Religious Aesthetics: A Theological Study of Making and Meaning.* Princeton: University Press, 1989. xvi, 225 p. Cloth $27.50.

BROWNING, DOUGLAS: *Ontology and the Practical Arena.* University Park: Penn State UP, 1990. xii, 167 p. Cloth $25.00.

BRUNER, JEROME: *Acts of Meaning.* Cambridge: Harvard, 1990. xvii, 179 p. Cloth $19.95.

BRYSON, NORMAN: *Looking at the Overlooked: Four Essays on Still Life Painting.* Cambridge, Harvard, 1990. 192 p. Cloth $25.00.

BUCHANAN, ALLEN E., and DAN W. BROCK: *Deciding for Others: The Ethics of Surrogate Decision Making.* New York: Cambridge, 1990. xv, 422 p. Cloth $49.50, paper $16.95.

BURROWS, DAVID: *Sound, Speech, and Music.* Amherst: Massachusetts UP, 1990. viii, 138 p. Cloth $20.00.

BYRNE, EDMUND F.: *Work, Inc.: A Philosophical Analysis.* Philadelphia: Temple, 1990. xi, 340 p. Cloth $39.95.

CAHN, STEVEN M.: *Philosophical Explorations: Freedom, God, and Goodness.* Buffalo: Prometheus, 1989. 231 p. Paper $14.95.

CALLINICOS, ALEX: *Against Postmodernism: A Marxist Critique.* New York: St. Martin's, 1990. x, 207 p. Cloth $49.95, paper $15.95.

CAMPBELL, KEITH: *Abstract Particulars.* Cambridge: Blackwell, 1990. xii, 187 p. Cloth $42.95.

CARRUTHERS, PETER: *The Metaphysics of the* Tractatus. New York: Cambridge, 1990. xiv, 210 p. Cloth $42.50; and *Tractarian Semantics: Finding Sense in Wittgenstein's* Tractatus. Cambridge: Blackwell, 1989. xii, 232 p. Cloth $45.00.

CASEY, JOHN: *Pagan Virtue: An Essay in Ethics.* New York: Oxford, 1990. xi, 242 p. Cloth $49.95.

CAYGILL, HOWARD: *Art of Judgement.* Cambridge: Blackwell, 1989. ix, 428 p. Cloth $39.95.

CHALMERS, ALAN: *Science and Its Fabrication.* Minneapolis: Minnesota UP, 1990. xii, 142 p. Cloth $39.95, paper $14.95.

CHATTOPADHYAYA, D. P.: *Anthropology and Historiography of Science.* Athens: Ohio UP, 1990. xxxi, 265 p. Cloth $34.95.

CHIERCHIA, GENNARO, and SALLY MCCONNELL-GINET: *Meaning and Grammar: An Introduction to Semantics.* Cambridge: MIT, 1990. xiii, 476 p. Cloth $29.95.

CHIHARA, CHARLES S.: *Constructibility and Mathematical Existence.* New York: Oxford, 1990. xv, 282 p. Cloth $39.95.

CLARK, KELLY JAMES: *Return to Reason.* Grand Rapids, MI: Eerdmans, 1990. ix, 158 p. Paper $10.95.

COBB-STEVENS, RICHARD: *Husserl and Analytic Philosophy.* Boston: Kluwer, 1990. vii, 222 p. Cloth $69.00.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

COBURN, ROBERT C.: *The Strangeness of the Ordinary: Problems and Issues in Contemporary Metaphysics.* Savage, MD: Rowman & Littlefield, 1990. xiii, 179 p. Cloth $34.50.

COLEMAN, JAMES S.: *Foundations of Social Theory.* Cambridge: Harvard, 1990. xvi, 993 p. Cloth $39.50.

CONDREN, CONAL: *George Lawson's* Politica *and the English Revolution.* New York: Cambridge, 1990. xviii, 211 p. Cloth $49.50.

CRUMP, THOMAS: *The Anthropology of Numbers.* New York: Cambridge, 1990. x, 201 p. Cloth $49.50.

DANFORD, JOHN W.: *David Hume and the Problem of Reason: Recovering the Human Sciences.* New Haven: Yale, 1990. xii, 228 p. Cloth $25.00.

DEELY, JOHN: *Basics of Semiotics.* Bloomington: Indiana UP, 1990. xvii, 150 p. Cloth $25.00, paper $9.95.

DONHAM, DONALD L.: *History, Power, Ideology: Central Issues in Marxism and Anthropology.* New York: Cambridge, 1990. xi, 242 p. Cloth $49.50.

DRUMMOND, JOHN J.: *Husserlian Intentionality and Non-Foundational Realism: Noema and Object.* Boston: Kluwer, 1990. xii, 295 p. Cloth $94.00.

DUNN, JOHN: *Interpreting Philosophical Responsibility.* Princeton: University Press, 1990. viii, 274 p. Cloth $49.50, paper $16.95.

EAGLETON, TERRY: *The Ideology of the Aesthetic.* Cambridge: Blackwell, 1990. v, 417 p. Cloth $49.95, paper $17.95.

EDEL, ABRAHAM: *The Struggle for Academic Democracy: Lessons from the 1938 "Revolution" in New York's City Colleges.* Philadelphia: Temple, 1990. viii, 285 p. Cloth $29.95.

EDWARDS, JAMES C.: *The Authority of Language: Heidegger, Wittgenstein, and the Threat of Philosophical Nihilism.* Tampa: South Florida UP, 1990. xi, 249 p. Cloth $28.95.

ELLIS, BRIAN: *Truth and Objectivity.* Cambridge: Blackwell, 1990. ix, 314 p. Cloth $44.95.

EPSTEIN, RICHARD L.: *The Semantic Foundations of Logic: Volume 1: Propositional Logics.* Boston: Kluwer, 1990. xxii, 386 p.

ETCHEMENDY, JOHN: *The Concept of Logical Consequence.* Cambridge: Harvard, 1990. xii, 174 p. Cloth $25.00.

FALES, EVAN: *Causation and Universals.* New York: Routledge, 1990. xxii, 362 p. Cloth $65.00.

FARMER, DAVID J.: *Being in Time: The Nature of Time in Light of Mctaggart's Paradox.* New York: UP of America, 1990. xi, 223 p. Cloth $33.50.

FERRAROTTI, FRANCO: *Time, Memory, and Society.* Westport, CT: Greenwood, 1990. x, 161 p. Cloth $39.95.

FODOR, JERRY A.: *A Theory of Content and Other Essays.* Cambridge: MIT, 1990. xii, 270 p. Cloth $29.95.

FORCE, JAMES E., and RICHARD H. POPKIN: *Essays on the Context, Nature, and Influence of Isaac Newton's Theology.* Boston: Kluwer, 1990. x, 226 p. Cloth $79.00.

FORRESTER, JOHN: *The Seductions of Psychoanalysis: Freud, Lacan, and Derrida.* New York: Cambridge, 1990. xii, 421 p. Cloth $49.50.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

FORSBERG, RALPH P.: *Thomas Hobbes' Theory of Obligation: A Modern Interpretation.* Wakefield, NH: Longwood, 1990. xvii, 267 p. Cloth $35.00.

FOX, MARVIN: *Interpreting Maimonides: Studies in Methodology, Metaphysics, and Moral Philosophy.* Chicago: University Press, 1990. xiii, 356 p. Cloth $27.00.

FRANCO, PAUL: *The Political Philosophy of Michael Oakeshott.* New Haven: Yale, 1990. vii, 277 p. Cloth $30.00.

FRANKLIN, ALLAN: *Experiment, Right or Wrong.* New York: Cambridge, 1990. x, 230 p. Cloth $44.50.

FROLOV, IVAN: *Man, Science, Humanism: A New Synthesis.* Buffalo: Prometheus, 1990. 342 p. Cloth $24.95.

FUMERTON, RICHARD A.: *Reason and Morality: A Defense of the Egocentric Perspective.* Ithaca: Cornell, 1990. xiii, 247 p. Cloth $29.95.

GAUS, GERALD F.: *Value and Justification: The Foundations of Liberal Theory.* New York: Cambridge, 1990. xviii, 540 p. Cloth $59.50, paper $19.95.

GAUTHIER, DAVID: *Moral Dealing: Contract, Ethics, and Reason.* Ithaca: Cornell, 1990. vi, 371 p. Cloth $43.95, paper $14.95.

GIBBARD, ALLAN: *Wise Choices, Apt Feelings: A Theory of Normative Judgment.* Cambridge: Harvard, 1990. x, 346 p. Cloth $35.00.

GILBERT, ALAN: *Democratic Individuality.* New York: Cambridge, 1990. xv, 510 p. Cloth $59.50, paper $19.95.

GILL, R. R. ROCKINGHAM: *Deducibility and Decidability.* New York: Routledge, 1990. viii, 162 p. Cloth $47.50.

GINET, CARL: *On Action.* New York: Cambridge, 1990. xi, 159 p. Cloth $34.50, paper $11.95.

GOLDSTEIN, LAURENCE: *The Philosopher's Habitat.* New York: Routledge, 1990. xvii, 215 p. Cloth $49.95, paper $14.95.

GOODMAN, NELSON: *A Study of Qualities.* New York: Garland, 1990 (1940). xiii, 657 p. Cloth $127.00.

GRAVE, S. A.: *Conscience in Newman's Thought.* New York: Oxford, 1989. vi, 191 p. Cloth $39.95.

GRAYBEAL, JEAN: *Language and 'the Feminine' in Nietzsche and Heidegger.* Bloomington: Indiana UP, 1990. vii, 182 p. Cloth $24.50, paper $10.95.

GROARKE, LEO: *Greek Scepticism: Anti-Realist Trends in Ancient Thought.* Montreal: McGill-Queen's UP, 1990. xv, 176 p. Cloth $34.95.

GROSSMANN, REINHARDT: *The Fourth Way: A Theory of Knowledge.* Bloomington: Indiana UP, 1990. viii, 312 p. Cloth $35.00.

HALE, MALCOLM I.: *The Mind: Its Origin, Evolution, Structure and Functioning.* Pittsburgh: Hale-van Ruth, 1989. xi, 224 p. Paper $7.95.

HAMLYN, D. W.: *In and Out of the Black Box: On the Philosophy of Cognition.* Cambridge: Blackwell, 1990. vii, 179 p. Cloth $29.95.

HARMAN, GILBERT: *Skepticism and the Definition of Knowledge.* New York: Garland, 1990. 163 p. Cloth $33.00.

HART, KEVIN: *The Trespass of the Sign: Deconstruction, Theology and Philosophy.* New York: Cambridge, 1989. xii, 292 p. Cloth $47.50.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

HARTSHORNE, CHARLES: *The Darkness and the Light: A Philosopher Reflects upon His Fortunate Career and Those Who Made It Possible.* New York: SUNY, 1990. xix, 426 p. Cloth $39.50.

HARTSHONE, M. HOLMES: *Kierkegaard, Godly Deceiver: The Nature and Meaning of His Pseudonymous Writings.* New York: Columbia, 1990. xvii, 112 p. Cloth $35.00.

HEAL, JANE: *Fact and Meaning: Quine and Wittgenstein on Philosophy of Language.* Cambridge: Blackwell, 1989. viii, 247 p. Cloth $39.95.

HEILKE, THOMAS W.: *Voegelin on the Idea of Race: An Analysis of Modern European Racism.* Baton Rouge: LSU UP, 1990. xiii, 161 p. Cloth $22.50.

HODGES, MICHAEL P.: *Transcendence and Wittgenstein's* Tractatus. Philadelphia: Temple, 1990. xii, 205 p. Cloth $39.95.

HOOK, SIDNEY: *Convictions.* Buffalo: Promethus, 1990. 310 p. Cloth $24.95.

HOPE, V. M.: *Virtue by Consensus: The Moral Philosophy of Hutchenson, Hume and Adam Smith.* New York: Oxford, 1989. viii, 166 p. Cloth $39.95.

HOWARD, DICK: *Defining the Political.* Minneapolis: Minnesota UP, 1989. xiii, 331 p. Cloth $39.50, paper $15.95; and *The Politics of Critique.* Minneapolis: Minnesota UP, 1988. xxviii, 245 p. Cloth $35.00, paper $14.95.

HUTTON, CHRISTOPHER M.: *Abstraction and Inference: The Type-Token Relation in Linguistics Theory.* New York: Pergamon, 1990. viii, 180 p. Cloth $52.95.

IHDE, DON: *Technology and the Lifeworld: From Garden to Earth.* Bloomington: Indiana UP, 1990. xiv, 226 p. Cloth $35.00, paper $14.95.

JOLIEY, NICHOLAS: *The Light of the Soul: Theories of Ideas in Leibniz, Malebranche, and Descartes.* New York: Oxford, 1990. x, 209 p. Cloth $39.95.

KENNY, ANTHONY: *The Metaphysics of Mind.* New York, Oxford, 1989. ix, 165 p. Cloth $24.95.

KERNAN, ALVIN: *The Death of Literature.* New Haven: Yale, 1990. ix, 230 p. Cloth $22.50.

KOLB, DAVID: *Postmodern Sophistications: Philosophy, Architecture, and Tradition.* Chicago: University Press, 1990. xi, 216 p. Cloth $29.95.

KOLENDA, KONSTANTIN: *Rorty's Humanistic Pragmatism: Philosophy Democratized.* Gainesville: Florida UP, 1990. xv, 138 p. Cloth $22.95.

KOVACS, GEORGE: *The Question of God in Heidegger's Phenomenology.* Evanston: Northwestern UP, 1990. xxi, 326 p. Cloth $39.95, paper $15.95.

KUKATHAS, CHANDRAN: *Hayek and Modern Liberalism.* New York: Oxford, 1989. xiv, 247 p. Cloth $39.95.

KUPPERS, BERND-OLAF: *Information and the Origin of Life.* Cambridge: MIT, 1990. xxiii, 215 p. Cloth $22.50.

KURTZ, PAUL: *Philosophical Essay in Pragmatic Naturalism.* Buffalo: Prometheus, 1990. 265 p. Cloth $19.95.

LANG, BEREL: *Act and Idea in the Nazi Genocide.* Chicago: University Press, 1990. xxii, 258 p. Cloth $34.95, paper $14.95; and *The Anatomy of*

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Philosophical Style. Cambridge: Blackwell, 1990. vii, 277 p. Cloth $44.95, paper $16.95.

LAWSON, E. THOMAS, and ROBERT N. McCAULEY: *Rethinking Religion: Connecting Cognition and Culture*. New York: Cambridge, 1990. ix, 194 p. Cloth $42.50.

LEAR, JONATHAN: *Love and its Place in Nature: A Philosophical Interpretation of Freudian Psychoanalysis*. New York: Farrar, Straus & Giroux, 1990. Cloth $17.95.

LEDER, DREW: *The Absent Body*. Chicago: University Press, 1990. x, 218 p. Cloth $34.95, paper $14.95.

LEDGER, GERARD R.: *Re-counting Plato: A Computer Analysis of Plato's Style*. New York: Oxford, 1990. xiv, 254 p. Cloth $68.00.

LINKLATER, ANDREW: *Beyond Realism and Marxism: Critical Theory and International Relations*. New York: St. Martin's, 1990. viii, 205 p. Cloth $45.00.

LLOYD, A. C.: *The Anatomy of Neoplatonism*. New York: Oxford, 1990. x, 198 p. Cloth $45.00.

LOCKWOOD, MICHAEL: *Mind, Brain, & the Quantum: The Compound 'I'*. Cambridge: Blackwell, 1990. xiii, 365 p. Cloth $29.95.

LOWE, E. JONATHAN: *Kinds of Being: A Study of Individuation, Identity and the Logic of Sortal Terms*. New York: Blackwell, 1989. vi, 210 p. Cloth $39.95.

McCARNEY, JOSEPH: *Social Theory and the Crisis of Marxism*. New York: Verso, 1990. x, 217 p. Cloth $50.00, paper $17.95.

McCLENNEN, EDWARD F.: *Rationality and Dynamic Choice: Foundational Explorations*. New York: Cambridge, 1990. xv, 311 p. Cloth $39.50.

McCORMICK, PETER J.: *Modernity, Aesthetics, and the Bounds of Art*. Ithaca: Cornell, 1990. xiii, 349 p. Cloth $42.50, paper $14.95.

McGEE, VANN: *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth*. Indianapolis: Hackett, 1990. x, 236 p. Cloth $28.50.

MCLAUGHLIN, ROBERT N.: *On the Logic of Ordinary Conditionals*. Albany: SUNY, 1990. xvii, 202 p. Cloth $54.50, paper $17.95.

MACHAN, TIBOR R., and JOHN O. NELSON: *A Dialogue Partly on Political Liberty*. New York: UP of America, 1990. 80 p. Paper $9.75.

MAKKREEL, RUDOLF A.: *Imagination and Interpretation in Kant: The Hermeneutical Import of the* Critique of Judgement. Chicago: University Press, 1990. xi, 187 p. Cloth $24.95.

MANKTELOW, K. I., and D. E. OVER: *Inference and Understanding: A Philosophical and Psychological Perspective*. New York: Routledge, 1990. vii, 194 p. Cloth $57.50, paper $16.50.

MARTIN, JAMES A., JR.: *Beauty and Holiness: The Dialogue between Aesthetics and Religion*. Princeton: University Press, 1990. ix, 222 p. Cloth $27.50.

MOHANTY, J. N.: *Transcendental Phenomenology*. Cambridge: Blackwell, 1990. viii, 176 p. Cloth $34.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

MOLES, ALISTAIR: *Nietzsche's Philosophy of Nature and Cosmology.* New York: Lang, 1990. xvii, 434 p. Cloth $63.95.

MOORE, A. W.: *The Infinite.* New York: Routledge, 1990. xiii, 268 p. Cloth $42.00.

MORAVCSIK, J. M.: *Thought and Language.* New York: Routledge, 1990. ix, 283 p. Cloth $35.00.

MORERA, ESTEVE: *Gramsci's Historicism: A Realist Interpretation.* New York: Routledge, 1990. vi, 237 p. Cloth $45.00.

MORGAN, MICHAEL L.: *Platonic Piety: Philosophy and Ritual in Fourth-Century Athens.* New Haven: Yale, 1990. x, 273 p. Cloth $28.50.

MOYA, CARLOS J.: *The Philosophy of Action.* Cambridge, UK: Polity, 1990. viii, 189 p. Cloth $49.95.

MULHALL, STEPHEN: *On Being in the World: Wittgenstein and Heidegger on Seeing Aspects.* New York: Routledge, 1990. ix, 206 p. Cloth $57.50.

MUNZER, STEPHEN R.: *A Theory of Property.* New York: Cambridge, 1990. x, 491 p. Cloth $55.00, paper $19.95.

MURPHEY, NANCY: *Theology in the Age of Scientific Reasoning.* Ithaca: Cornell, 1990. xii, 215 p. Cloth $26.95.

NEALE, STEPHEN: *Descriptions.* Cambridge: MIT, 1990. xiv, 286 p. Cloth $29.95.

NELSON, LYNN HANKINSON: *Who Knows: From Quine to a Feminist Empiricism.* Philadelphia: Temple, 1990. xi, 401 p. Cloth $34.95.

ORTONY, ANDREW, GERARD L. CLORE, and ALLAN COLLINS: *The Cognitive Structure of Emotions.* New York: Cambridge, 1990. xi, 207 p. Paper $14.95.

OSTROW, JAMES M.: *Social Sensitivity: A Study of Habit and Experience.* Albany: SUNY, 1990. viii, 137 p. Cloth $34.50, paper $10.95.

PEFFER, R. G.: *Marxism, Morality, and Social Justice.* Princeton: University Press, 1990. xiv, 526 p. Cloth $49.50, paper $16.95.

PELLETIER, FRANCIS JEFFREY: *Parmenides, Plato, and the Semantics of Not-Being.* Chicago: University Press, 1990. xxi, 166 p. Cloth $29.95.

PIVČEVIĆ, EDO: *Change and Selves.* New York: Oxford, 1990. x, 151 p. Cloth $39.95.

POLKA, BRAYTON: *Truth and Interpretation: An Essay in Thinking.* New York: St. Martin's, 1990. xxxvi, 340 p. Cloth $35.00.

POLLARD, STEPHEN: *Philosophical Introduction to Set Theory.* Notre Dame: University Press, 1990. xii, 180 p. Cloth $24.95, paper $12.95.

POLLOCK, JOHN L.: *Nomic Probability and the Foundations of Induction.* New York: Oxford, 1990. xviii, 344 p. Cloth $32.50; and *Technical Methods in Philosophy.* Boulder: Westview, 1990. x, 126 p. Cloth $33.95, paper $11.95.

POMPA, LEON: *Human Nature and Historical Knowledge: Hume, Hegel and Vico.* New York: Cambridge, 1990. viii, 234 p. Cloth $44.50.

POSNER, RICHARD: *The Problems of Jurisprudence.* Cambridge: Harvard, 1990. xiv, 485 p. Cloth $29.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

PUTNAM, HILARY: *The Concept of Probability in Application to Finite Sequences.* New York: Garland, 1990 (1951). xiv, 128 p. Cloth $22.00; and *Realism with a Human Face.* Cambridge: Harvard, 1990. lxxiv, 347 p. Cloth $30.00.

QUINE, W. V.: *The Logic of Sequences: A Generalization of* Principia Mathematica. New York: Garland, 1990 (1932). iv, 290 p. Cloth $57.00.

RAMBERG, BJORN T.: *Donald Davidson's Philosophy of Language: An Introduction.* Cambridge: Blackwell, 1989. iv, 153 p. Cloth $39.95, paper $15.95.

RASMUSSEN, DAVID M.: *Reading Habermas.* Cambridge: Blackwell, 1990. vii, 146 p. Paper $17.95.

RENTELN, ALISON DUNDES: *International Human Rights: Universalism versus Relativism.* Newbury Park, CA: Sage, 1990. 205 p. Cloth $35.00, paper $16.95.

RESCHER, NICHOLAS: *A Useful Inheritance: Evolutionary Aspects of the Theory of Knowledge.* Savage, MD: Rowman & Littlefield, 1990. x, 138 p. Cloth $28.50.

RICHARD, MARK: *Propositional Attitudes: An Essay on Thoughts and How We Ascribe Them.* New York: Cambridge, 1990. ix, 275 p. Cloth $44.50, paper $16.95.

RITVO, LUCILLE B.: *Darwin's Influence on Freud: A Tale of Two Sciences.* New Haven: Yale, 1990. xii, 267 p. Cloth $30.00.

ROOSEVELT, GRACE G.: *Reading Rousseau in the Nuclear Age.* Philadelphia: Temple, 1990. x, 275 p. Cloth $44.95.

RUCHLIS, HY with SANDRA ODDO: *Clear Thinking: A Practical Introduction.* Buffalo: Prometheus, 1990. x, 271 p. Paper $15.95.

RUNDLE, BEDE: *Wittgenstein and Contemporary Philosophy of Language.* Cambridge: Blackwell, 1990. x, 274 p. Cloth $49.95.

SALKEVER, STEPHEN G.: *Finding the Mean: Theory and Practice in Aristotelian Political Philosophy.* Princeton: University Press, 1990. x, 287 p. Cloth $35.00.

SALLIS, JOHN: *Echoes: After Heidegger.* Bloomington: Indiana UP, 1990. ix, 213 p. Cloth $27.50, paper $12.50.

SCHAEFFER, JOHN D.: *Sensus Communis: Vico, Rhetoric and the Limits of Relativism.* Durham, NC: Duke, 1990. x, 182 p. Cloth $25.75.

SCHWEIKER, WILLIAM: *Mimetic Reflections: A Study in Hermeneutics, Theology, and Ethics.* New York: Fordham, 1990. xiii, 267 p. Cloth $40.00, paper $19.95.

SCHWYZER, HUBERT: *The Unity of Understanding: A Study in Kantian Problems.* New York: Oxford, 1990. ix, 172 p. Cloth $26.00.

SCOTT, CHARLES E.: *The Question of Ethics: Nietzsche, Foucault, Heidegger.* Bloomington: Indiana UP, 1990. xii, 225 p. Cloth $29.95, paper $12.95.

SCOTT, G. E.: *Moral Personhood: An Essay in the Philosophy of Moral Psychology.* Albany: SUNY, 1990. xvi, 202 p. Cloth $44.50, paper $14.95.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

SHAPIRO, BRUCE G.: *Divine Madness and the Absurd Paradox: Ibsen's* Peer Gynt *and the Philosophy of Kierkegaard.* Westport, CT: Greenwood, 1990. xxi, 247 p. Cloth $39.95.

SHAPIRO, IAN: *Political Criticism.* Berkeley: California UP, 1990. xii, 338 p. Cloth $35.00.

SHKLAR, JUDITH N.: *The Faces of Injustice.* New Haven: Yale, 1990. vii, 144 p. Cloth $19.95.

SKYRMS, BRIAN: *The Dynamics of Rational Deliberation.* Cambridge: Harvard, 1990. x, 199 p. Cloth $32.50.

SLEIGH, R. C., JR.: *Leibniz and Arnauld: A Commentary on Their Correspondence.* New Haven: Yale, 1990. xv, 237 p. Cloth $28.50.

SMART, J. J. C.: *Our Place in the Universe: A Metaphysical Discussion.* New York: Blackwell, 1989. x, 212 p. Cloth $39.95.

SMITH, TONY: *The Logic of Marx's* Capital: *Replies to Hegelian Criticisms.* Albany: SUNY, 1990. xii, 271 p. Cloth $54.50, paper $17.95.

SOBLE, ALAN: *The Structure of Love.* New Haven: Yale, 1990. xiv, 374 p. Cloth $32.50.

STERN, ROBERT: *Hegel, Kant and the Structure of the Object.* New York: Routledge, 1990. xiii, 169 p. Cloth $54.00.

STERNBERG, ROBERT J.: *Wisdom: Its Nature, Origins, and Development.* New York: Cambridge, 1990. x, 339 p. Cloth $42.50, paper $16.95.

STICH, STEPHEN: *The Fragmentation of Reason.* Cambridge: MIT, 1990. x, 181 p. Cloth $22.50.

STOCKER, MICHAEL: *Plural and Conflicting Values.* New York: Oxford, 1990. xiii, 360 p. Cloth $39.95.

TAMBIAH, STANLEY J.: *Magic, Science, Religion, and the Scope of Rationality.* New York: Cambridge, 1990. xi, 187 p. Cloth $42.50, paper $13.95.

TOWNSEND, DABNEY: *Aesthetic Objects and Works of Art.* Wolfeboro, NH: Longwood, 1990. ix, 299 p. Cloth $30.00, paper $14.95.

UNGER, PETER: *Identity, Consciousness, and Value.* New York: Oxford, 1990. xiv, 344 p. Cloth $38.00.

VANHOOZER, KEVIN J.: *Biblical Narrative in the Philosophy of Paul Ricoeur: A Study in Hermeneutics and Theology.* New York: Cambridge, 1990. xiii, 308 p. Cloth $49.50.

WALLACH, MICHAEL A., and LISE WALLACH: *Rethinking Goodness.* New York: SUNY, 1990. x, 156 p. Cloth $44.50, paper $14.95.

WALTON, DOUGLAS N.: *Practical Reasoning: Goal-Driven, Knowledge-Based, Action-Guiding Argumentation.* Savage, MD: Rowman & Littlefield, 1990. Cloth $38.50.

WALTON, KENDALL L.: *Mimesis as Make-Believe: On the Foundations of the Representational Arts.* Cambridge: Harvard, 1990. xvii, 450 p. Cloth $35.00.

WEINER, JOAN: *Frege in Perspective.* Ithaca: Cornell, 1990. xvii, 307 p. Cloth $32.50.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

INDEX OF SUBJECTS DISCUSSED IN THE JOURNAL

Abortion

Cudd, Ann E.: Sensationalized Philosophy: A Reply to Marquis's "Why Abortion Is Immoral." **87** (5): 262–4 '90

McInerney, Peter K.: Does a Fetus Already Have a Future-like-ours? **87** (5): 264–8 '90

Norcross, Alastair: Killing, Abortion, and Contraception: A Reply to Marquis. **87** (5): 268–77 '90

Aesthetics

Goldman, Alan H.: Aesthetic Qualities and Aesthetic Value. **87** (1): 23–37 '90

American Philosophical Association, Eastern Division

The 87th Annual Meeting. [Program.] **87** (10): 551–84 '90; [Symposium Papers.] **87** (10, 11): 493–547, 585–668 '90

Anaphora, Donkey

Neale, Stephen: Descriptive Pronouns and Donkey Anaphora. **87** (3): 113–50 '90

Appiah, Anthony

Wartenberg, Thomas E.: Comments on Appiah and Lugones. [Abstract.] **87** (10): 508–9 '90

Argument

Walton, Douglas N.: What Is Reasoning? What Is an Argument? **87** (8): 399–419 '90

Aune, Bruce

Rosenberg, Jay F.: Response to Aune. [Abstract.] **87** (10): 546–7 '90

Benevolence

Estlund, David M.: Mutual Benevolence and the Theory of Happiness. **87** (4): 187–204 '90

Scoccia, Danny: Utilitarianism, Sociobiology, and the Limits of Benevolence. **87** (7): 329–45 '90

Choice

Levi, Isaac: Pareto Unanimity and Consensus. **87** (9): 481–92 '90

Church, Alonzo

Mendelson, Elliott: Second Thoughts about Church's Thesis and Mathematical Proofs. **87** (5): 225–33 '90

Cognitivism, Moral

Schiffer, Stephen: Meaning and Value. **87** (11): 602–14 '90

Sturgeon, Nicholas: Schiffer on Meaning and Value. [Abstract.] **87** (11): 615–6 '90

Wong, David: A Relativist Alternative to Antirealism. [Abstract.] **87** (11): 617–8 '90

Colapietro, Vincent M.

Stuhr, John J.: Subjects Constructed, Deconstructed, and Reconstructed. **87** (11): 262–4 '90

Commodities

Bacharach, Michael: Commodities, Language, and Desire. **87** (7): 346–68 '90

Conditionals

Thomson, James F.: In Defense of '⊃'. **87** (2): 57–70 '90

Consensus

Levi, Isaac: Pareto Unanimity and Consensus. **87** (9): 481–92 '90

Descriptions

Neale, Stephen: Descriptive Pronouns and Donkey Anaphora. **87** (3): 113–50 '90

Desire

Bacharach, Michael: Commodities, Language, and Desire. **87** (7): 346–68 '90

Dewey, John

Capaldi, Nicholas: Hook, Dewey, and

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Marx. [Abstract.] **87** (10): 535–6 '90
Davidson, Donald: The Dewey Lectures 1989: The Structure and Content of Truth. **87** (6): 279–328 '90

Dilemma, Prisoner's

Blais, Michel J.: Misunderstandings of Epistemic TIT for TAT: Reply to John Woods. **87** (7): 369–74 '90

Discourse

Colapietro, Vincent M.: The Vanishing Subject of Contemporary Discourse: A Pragmatic Response. **87** (11): 644–55 '90
Stuhr, John J.: Subjects Constructed, Deconstructed, and Reconstructed. [Abstract.] **87** (11): 656–7 '90

Driver, Julia

Flanagan, Owen: Virtue and Ignorance. **87** (8): 420–8 '90

Essentialism

Steward, Helen: Identity Statements and the Necessary A Posteriori. **87** (8): 385–98 '90

Ethnicity

Appiah, Anthony: But Would That Still Be Me? Notes on Gender, "Race," Ethnicity, as Sources of "Identity." **87** (10): 493–9 '90
Lugones, Maria C.: Structure/Antistructure and Agency under Oppression. **87** (10): 500–7 '90
Wartenberg, Thomas E.: Comments on Appiah and Lugones. [Abstract.] **87** (10): 508–9 '90

Explanation

Vogel, Jonathan: Cartesian Skepticism and Inference to the Best Explanation. **87** (11): 658–66 '90

Famine

Aiken, William: Famine and Distribution. [Abstract.] **87** (11): 642–3 '90
Lucas, George R., Jr.: African Famine: New Economic and Ethical Perspectives. **87** (11): 629–41 '90

Freedom, Political

Tully, James: Political Freedom. **87** (10): 517–23 '90

Freud, Sigmund

Myers, Gerald E.: James and Freud. **87** (11): 593–9 '90

Gender

Appiah, Anthony: But Would That Still Be Me? Notes on Gender, "Race," Ethnicity, as Sources of "Identity." **87** (10): 493–9 '90
Lugones, Maria C.: Structure/Antistructure and Agency under Oppression. **87** (10): 500–7 '90
Wartenberg, Thomas E.: Comments on Appiah and Lugones. [Abstract.] **87** (10): 508–9 '90

Happiness

Estlund, David M.: Mutual Benovolence and the Theory of Happiness. **87** (4): 187–204 '90

Hook, Sydney

Capaldi, Nicholas: Hook, Dewey, and Marx. [Abstract.] **87** (10): 535–6 '90
Kurtz, Paul: Pragmatic Naturalism: The Philosophy of Sidney Hook (1902–1989). **87** (10): 526–34 '90

Identity

Steward, Helen: Identity Statements and the Necessary A Posteriori. **87** (8): 385–98 '90

Identity, Personal

Appiah, Anthony: But Would That Still Be Me? Notes on Gender, "Race," Ethnicity, as Sources of "Identity." **87** (10): 493–9 '90
Lugones, Maria C.: Structure/Antistructure and Agency under Oppression. **87** (10): 500–7 '90
Schechtman, Marya: Personhood and Personal Identity. **87** (2): 71–92 '90
Wartenberg, Thomas E.: Comments on Appiah and Lugones. [Abstract.] **87** (10): 508–9 '90

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

James, William

Myers, Gerald E.: James and Freud. **87** (11): 593–9 '90

Seigfried, Charlene Haddock: The Pragmatist Sieve of Concepts: Description versus Interpretation. **87** (11): 585–92 '90

Thayer, H. S.: Comments on Seigfried and Myers. [Abstract.] **87** (11): 600–1 '90

Kitcher, Philip

Sober, Elliott: The Poverty of Pluralism. **87** (3): 151–8 '90

Kurtz, Paul

Capaldi, Nicholas: Hook, Dewey, and Marx. [Abstract.] **87** (10): 535–6 '90

Language, Philosophy of

Bacharach, Michael: Commodities, Language, and Desire. **87** (7): 346–68 '90

Neale, Stephen: Descriptive Pronouns and Donkey Anaphora. **87** (3): 113–50 '90

Linguistics

Neale, Stephen: Descriptive Pronouns and Donkey Anaphora. **87** (3): 113–50 '90

Locke, John

Chappell, Vere C.: Locke's Moral Psychology. [Abstract.] **87** (10): 524–5 '90

Tully, James: Political Freedom. **87** (10): 517–23 '90

Yolton, John W.: The Way of Ideas: A Retrospective. **87** (10): 510–6 '90

Logic, Philosophy of

Thomson, James F.: In Defense of '⊃'. **87** (2): 57–70 '90

Logic, Second-order

Shapiro, Stewart: Second-order Logic, Foundations, and Rules. **87** (5): 234–61 '90

Lucas, George R., Jr.

Aiken, William: Famine and Distribution. [Abstract.] **87** (11): 642–3 '90

Lugones, Maria C.

Wartenberg, Thomas E.: Comments on Appiah and Lugones. [Abstract.] **87** (10): 500–7 '90

Marquis, Don

Cudd, Ann E.: Sensationalized Philosophy: A Reply to Marquis's "Why Abortion Is Immoral." **87** (5): 262–4 '90

McInerney, Peter K.: Does a Fetus Already Have a Future-like-ours? **87** (5): 264–8 '90

Norcross, Alastair: Killing, Abortion, and Contraception: A Reply to Marquis. **87** (5): 268–77 '90

Marx, Karl

Capaldi, Nicholas: Hook, Dewey, and Marx. [Abstract.] **87** (10): 535–6 '90

Mathematics, Philosophy of

Maddy, Penelope: A Problem in the Foundations of Set Theory. **87** (11): 619–28 '90

Mendelson, Elliott: Second Thoughts about Church's Thesis and Mathematical Proofs. **87** (5): 225–33 '90

Shapiro, Stewart: Second-order Logic, Foundations, and Rules. **87** (5): 234–61 '90

Monism

Kitcher, Philip, Kim Sterelny, and C. Kenneth Waters: Illusory Riches of Sober's Monism. **87** (3): 158–61 '90

Motivation

Morillo, Carolyn R.: The Reward Event and Motivation. **87** (4): 169–86 '90

Myers, Gerald E.

Thayer, H. S.: Comments on Seigfried and Myers. [Abstract.] **87** (11): 600–1 '90

Naturalism

Kurtz, Paul: Pragmatic Naturalism: The Philosophy of Sidney Hook (1902–1989). **87** (10): 526–34 '90

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

New Books.

87 (4): 217–24, **87** (12): 721–9 '90

New Books: Anthologies.

87 (1): 48–56 '90

New Books: Translations.

87 (8): 437–8, '90

Notes and News.

87 (3): 165–7, **87** (5): 278, **87** (8): 436–7, **87** (10): 549–50, **87** (11): 669–70 '90

Ockhamism

Widerker, David: Troubles with Ockhamism. **87** (9): 462–80 '90

Pareto

Levi, Isaac: Pareto Unanimity and Consensus. **87** (9): 481–92 '90

Parfit, Derek

Schechtman, Marya: Personhood and Personal Identity. **87** (2): 71–92 '90

Philosophy, Continental

Colapietro, Vincent M.: The Vanishing Subject of Contemporary Discourse: A Pragmatic Response. **87** (11): 644–55 '90

Stuhr, John J.: Subjects Constructed, Deconstructed, and Reconstructed. [Abstract.] **87** (11): 656–7 '90

Pornography

Parent, W. A.: A Second Look at Pornography and the Subordination of Women. **87** (4): 205–11 '90

Psychoanalysis

Myers, Gerald E.: James and Freud. **87** (11): 593–9 '90

Psychology

Chappell, Vere C.: Locke's Moral Psychology. **87** (10): 524–5 '90

Myers, Gerald E.: James and Freud. **87** (11): 593–9 '90

Seigfried, Charlene Haddock: The Pragmatist Sieve of Concepts: Description versus Interpretation. **87** (11): 585–92 '90

Qualities, Aesthetic

Goldman, Alan H.: Aesthetic Qualities and Aesthetic Value. **87** (1): 23–37 '90

Race

Appiah, Anthony: But Would That Still Be Me? Notes on Gender, "Race," Ethnicity, as Sources of "Identity." **87** (10): 493–9 '90

Lugones, Maria C.: Structure/Antistructure and Agency under Oppression. **87** (10): 500–7 '90

Wartenberg, Thomas E.: Comments on Appiah and Lugones. [Abstract.] **87** (10): 508–9 '90

Realism, Moral

Schiffer, Stephen: Meaning and Value. **87** (11): 602–14 '90

Sturgeon, Nicholas: Schiffer on Meaning and Value. **87** (11): 615–6 '90

Wong, David: A Relativist Alternative to Antirealism. [Abstract.] **87** (11): 617–8 '90

Reasoning

Walton, Douglas N.: What Is Reasoning? What is an Argument? **87** (8): 399–419 '90

Relativism

Wong, David: A Relativist Alternative to Antirealism. [Abstract.] **87** (11): 617–8 '90

Schiffer, Stephen

Sturgeon, Nicholas: Schiffer on Meaning and Value. [Abstract.] **87** (11): 615–6 '90

Wong, David: A Relativist Alternative to Antirealism. [Abstract.] **87** (11): 617–8 '90

Science, Philosophy of

Kitcher, Philip: The Division of Cognitive Labor. **87** (1): 5–22 '90

Sober, Elliott: The Poverty of Pluralism. **87** (3): 151–8 '90

Seidenfield, Teddy

Levi, Isaac: Pareto Unanimity and Consensus. **87** (9): 481–92 '90

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Seigfried, Charlene Haddock

Thayer, H. S.: Comments on Seigfried and Myers. [Abstract.] **87** (11): 585–92 '90

Self-determination, National

Margalit, Avishai, and Joseph Raz: National Self-determination. **87** (9): 439–61 '90

Sellars, Wilfrid

Aune, Bruce: Sellar's Two Images of the World. **87** (10): 537–45 '90

Rosenberg, Jay F.: Response to Aune. [Abstract.] **87** (10): 546–7 '90

Set Theory

Maddy, Penelope: A Problem in the Foundations of Set Theory. **87** (11): 619–28 '90

Skepticism

Enc, Berent: Is Realism Really the Best Hypothesis? [Abstract.] **87** (11): 667–8 '90

Vogel, Jonathan: Cartesian Skepticism and Inference to the Best Explanation. **87** (11): 658–66 '90

Sober, Elliott

Kitcher, Philip, Kim Sterelny, and C. Kenneth Waters: Illusory Riches of Sober's Monism. **87** (3): 158–61 '90

Sociobiology

Scoccia, Danny: Utilitarianism, Sociobiology, and the Limits of Benevolence. **87** (7): 329–45 '90

Sterelny, Kim

Sober, Elliott: The Poverty of Pluralism. **87** (3): 151–8 '90

Subject

Colapietro, Vincent M.: The Vanishing Subject of Contemporary Discourse: A Pragmatic Response. **87** (11): 644–55 '90

Stuhr, John J.: Subjects Constructed, Deconstructed, and Reconstructed. [Abstract.] **87** (11): 656–7 '90

Truth

Davidson, Donald: The Dewey Lectures 1989: The Structure and Content of Truth. **87** (6): 279–328 '90

Utilitarianism

Scoccia, Danny: Utilitarianism, Sociobiology, and the Limits of Benevolence. **87** (7): 329–45 '90

Vadas, Melinda

Parent, W. A.: A Second Look at Pornography and the Subordination of Women. **87** (4): 205–11 '90

Value, Theory of

Goldman, Alan H.: Aesthetic Qualities and Aesthetic Value. **87** (1): 23–37 '90

Virtue

Flanagan, Owen: Virtue and Ignorance. **87** (8): 420–8 '90

Vogel, Jonathan

Enc, Berent: Is Realism Really the Best Hypothesis? [Abstract.] **87** (11): 658–66 '90

Women

Parent, W. A.: A Second Look at Pornography and the Subordination of Women. **87** (4): 205–11 '90

Woods, John

Blais, Michel J.: Misunderstandings of Epistemic TIT for TAT: Reply to John Woods. **87** (7): 369–74 '90

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

INDEX OF AUTHORS

Including Contributing Authors and Reviewers

Adler, Jonathan, Review of Nelson Goodman and Catherine Z. Elgin: *Reconceptions in Philosophy and Other Arts and Sciences.* **87** (12): 711–6 '90

Aiken, William, Famine and Distribution. [Abstract.] **87** (11): 642–3 '90

Akins, Kathleen A., Review of Patricia Churchland: *Neurophilosophy: Toward a Unified Theory of the Mind/Brain.* **87** (2): 93–102 '90

Appiah, Anthony, But Would That Still Be Me? Notes on Gender, "Race," Ethnicity, as Sources of "Identity." **87** (10): 493–9 '90

Aune, Bruce, Sellars's Two Images of the World. **87** (10): 537–45 '90

Bacharach, Michael, Commodities, Language, and Desire. **87** (7): 346–68 '90

Blais, Michel J., Misunderstandings of Epistemic TIT for TAT: Reply to John Woods. **87** (7): 369–74 '90

Capaldi, Nicholas, Hook, Dewey, and Marx. [Abstract.] **87** (10): 535–6 '90

Chappell, Vere C., Locke's Moral Psychology. [Abstract.] **87** (10): 524–5 '90

Cohen, Ted, Review of Mary Mothersill: *Beauty Restored.* **87** (12): 702–8 '90

Colapietro, Vincent M., The Vanishing Subject of Contemporary Discourse: A Pragmatic Response. **87** (12): 644–55 '90

Cudd, Ann E., Sensationalized Philosophy: A Reply to Marquis's "Why Abortion Is Immoral." **87** (5): 262–4 '90.

Davidson, Donald, The Dewey Lectures 1989: The Structure and Content of Truth. **87** (6): 279–328 '90

Enc, Berent, Is Realism Really the Best Hypothesis? [Abstract.] **87** (11): 667–8 '90

Estlund, David M., Mutual Benevolence and the Theory of Happiness. **87** (4): 187–204 '90

Flanagan, Owen, Virtue and Ignorance. **87** (8): 420–8 '90

Gaukroger, Stephen, Review of Edward Craig: *The Mind of God and the Works of Man.* **87** (8): 429–32 '90

Gert, Heather, J., Rights and Rights Violators: A New Approach to the Nature of Rights. **87** (12): 688–94 '90

Goldman, Alan, Aesthetic Qualities and Aesthetic Value. **87** (1): 23–37 '90

Grover, Dorothy, Truth and Language-world Connections. **87** (12): 671–87 '90

Hacking, Ian, Review of Peter Galison: *How Experiments End.* **87** (2): 103–6 '90

Hill, Christopher S., Review of Philip Pettit and John McDowell, eds.: *Subject, Thought, and Context.* **87** (2): 106–12 '90

Joy, Lynn S., Review of Marjorie Grene: *Descartes.* **87** (3): 161–5 '90

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Kitcher, Philip

The Division of Cognitive Labor. **87** (1): 5–22 '90

The Illusory Riches of Sober's Monism. By Philip Kitcher, Kim Sterelny, and C. Kenneth Waters. **87** (3): 158–61 '90

Kurtz, Paul, Pragmatic Naturalism: The Philosophy of Sidney Hook (1902–1989). **87** (10): 526–34 '90

Levi, Isaac, Pareto Unanimity and Consensus. **87** (9): 481–92 '90

Lucas, George R. Jr., African Famine: New Economic and Ethical Perspectives. **87** (11): 629–41 '90

Lugones, Maria C., Structure/Antistructure and Agency under Oppression. **87** (10): 500–7 '90

MacIntyre, Alasdair, Review of Richard Rorty: *Contingency, Irony, and Solidarity.* **87** (12): 708–11 '90

McInerney, Peter K., Does a Fetus Already Have a Future-like-ours? **87** (5): 264–8 '90

Maddy, Penelope, A Problem in the Foundations of Set Theory. **87** (11): 619–28 '90

Margalit, Avishai, National Self-determination. By Avishai Margalit and Joseph Raz. **87** (9): 439–61 '90

Mendelson, Elliott, Second Thoughts about Church's Thesis and Mathematical Proofs. **87** (5): 225–33 '90

Miller, Richard B., Supervenience Is a Two-way Street. **87** (12): 695–701 '90

Morillo, Carolyn R., The Reward Event and Motivation. **87** (4): 169–86 '90

Myers, Gerald E., James and Freud. **87** (11): 593–9 '90

Neale, Stephen, Descriptive Pronouns and Donkey Anaphora. **87** (3): 113–50 '90

Norcross, Alastair, Killing, Abortion, and Contraception: A Reply to Marquis. **87** (5): 268–77 '90

Parent, W. A., A Second Look at Pornography and the Subordination of Women. **87** (4): 205–11 '90

Pogge, Thomas W., Review of Brian Barry: *A Treatise of Social Justice, Vol. I: Theories of Justice.* **87** (7): 375–84 '90

Raz, Joseph, National Self-determination. By Avishai Margalit and Joseph Raz. **87** (9): 439–61 '90

Rosenberg, Jay F., Response to Aune. [Abstract.] **87** (10): 546–7 '90

Schechtman, Marya, Personhood and Personal Identity. **87** (2): 71–92 '90

Schiffer, Stephen, Meaning and Value. **87** (11): 602–14 '90

Scoccia, Danny, Utilitarianism, Sociobiology, and the Limits of Benevolence. **87** (7): 329–45 '90

Seigfried, Charlene Haddock, The Pragmatist Sieve of Concepts: Description versus Interpretation. **87** (11): 585–92 '90

Shapiro, Stewart, Second-order Logic, Foundations, and Rules. **87** (5): 234–61 '90

Shoemaker, Sydney, Review of Daniel Dennett: *The Intentional Stance.* **87** (4): 212–6 '90

Sklar, Lawrence, Review of Norman Swartz: *The Concept of Physical Law.* **87** (8): 432–5 '90

Sober, Elliott, The Poverty of Pluralism. **87** (3): 151–8 '90

Sterelny, Kim, The Illusory Riches of Sober's Monism. By Philip Kitcher, Kim Sterelny, and C. Kenneth Waters. **87** (3): 158–61 '90

Steward, Helen, Identity Statements and the Necessary A Posteriori. **87** (8): 385–98 '90

Stuhr, John J., Subjects Constructed, Deconstructed, and Reconstructed. [Abstract.] **87** (11): 656–7 '90

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

Sturgeon, Nicholas L., Schiffer on Meaning and Value. [Abstract.] **87** (11): 615–6 '90

Taschek, William W., Review of Kent Bach: *Thought and Reference.* **87** (1): 38–45 '90

Thayer, H. S., Comments on Seigfried and Myers. [Abstract.] **87** (11): 600–1 '90

Thomson, James F., In Defense of '⊃'. **87** (2): 57–70 '90

Tully, James, Political Freedom. **87** (10): 517–23 '90

Vogel, Jonathan, Cartesian Skepticism and Inference to the Best Explanation. **87** (11): 658–66 '90

Walton, Douglas N., What Is Reasoning? What Is an Argument? **87** (8): 399–419 '90

Wartenberg, Thomas E., Comments on Appiah and Lugones. [Abstract.] **87** (10): 508–9 '90

Waters, C. Kenneth, The Illusory Riches of Sober's Monism. By Philip Kitcher, Kim Sterelny, and C. Kenneth Waters. **87** (3): 158–61 '90

Widerker, David, Troubles with Ockhamism. **87** (9): 462–80 '90

Wolff, Robert Paul, Review of Thomas W. Pogge: *Realizing Rawls.* **87** (12): 716–20 '90

Wong, David B., A Relativist Alternative to Antirealism. [Abstract.] **87** (11): 617–8 '90

Yolton, John W., The Way of Ideas: A Retrospective. **87** (10): 510–6 '90

INDEX OF BOOKS REVIEWED

Listed by Author of Item Reviewed

Bach, Kent: *Thought and Reference.*—William Taschek. **87** (1): 38–45 '90

Barry, Brian: *A Treatise of Social Justice, Vol. I: Theories of Justice.*—Thomas W. Pogge. **87** (7): 375–84 '90

Churchland, Patricia: *Neurophilosophy: Toward a Unified Theory of the Mind/Brain.*—Kathleen A. Akins. **87** (2): 93–102 '90

Craig, Edward: *The Mind of God and the Works of Man.*—Stephen Gaukroger. **87** (8): 429–32 '90

Dennett, Daniel: *The Intentional Stance.*—Sydney Shoemaker. **87** (4): 212–6 '90

Galison, Peter: *How Experiments End.*—Ian Hacking. **87** (2): 103–6 '90

Goodman, Nelson, and Catherine Elgin: *Reconceptions in Philosophy and Other Arts and Sciences.*—Jonathan Adler. **87** (12): 711–6 '90

Grene, Marjorie: *Descartes.*—Lynn S. Joy. **87** (3): 161–5 '90

McDowell, John: *Subject, Thought, and Context.* By John McDowell and Christopher Pettit.—Christopher B. Hill. **87** (2): 106–12 '90

Mothersill, Mary: *Beauty Restored.*—Ted Cohen. **87** (12): 702–8 '90

Pettit, Christopher: *Subject, Thought, and Context.* By John McDowell and Christopher Pettit.—Christopher S. Hill. **87** (2): 106–12 '90

Pogge, Thomas W.: *Realizing Rawls.*—Robert Paul Wolff. **87** (12): 716–20 '90

Rorty, Richard: *Contingency, Irony, and Solidarity.*—Alasdair MacIntyre. **87** (12): 708–11 '90

Swartz, Norman: *The Concept of Physical Law.*—Lawrence Sklar. **87** (8): 432–5 '90

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION

Required by 39 U.S.C. 3685)

1A. Title of Publication	1B. PUBLICATION NO.	2. Date of Filing
THE JOURNAL OF PHILOSOPHY	0 0 2 2 3 6 2 X	9/13/90

3. Frequency of Issue	3A. No. of Issues Published Annually	3B. Annual Subscription Price
MONTHLY	12	$50. $30. $20

4. Complete Mailing Address of Known Office of Publication *(Street, City, County, State and ZIP+4 Code) (Not printer)*

709 Philosophy Hall, Columbia University, New York, New York 10027

5. Complete Mailing Address of the Headquarters of General Business Offices of the Publisher *(Not printer)*

709 Philosophy Hall, Columbia University, New York, New York 10027

6. Full Names and Complete Mailing Address of Publisher, Editor, and Managing Editor *(This item MUST NOT be blank)*

Publisher *(Name and Complete Mailing Address)*

THE JOURNAL OF PHILOSOPHY INC.
709 Philosophy Hall, Columbia University, New York, New York 10027

Editor *(Name and Complete Mailing Address)*

Arthur Danto, 709 Philosophy Hall, New York, New York 10027

Managing Editor *(Name and Complete Mailing Address)*

Michael Kelly, 704 Philosophy Hall, New York, NY 10027

7. Owner *(If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given. If the publication is published by a nonprofit organization, its name and address must be stated.) (Item must be completed.)*

Full Name	Complete Mailing Address
The Journal of Philosophy Inc.	709 Philosophy Hall, Columbia University New York, New York 10027

8. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages or Other Securities *(If there are none, so state)*

Full Name	Complete Mailing Address
None	

9. For Completion by Nonprofit Organizations Authorized To Mail at Special Rates *(DMM Section 423.12 only)*
The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes *(Check one)*

(1) [x] Has Not Changed During Preceding 12 Months
(2) [ ] Has Changed During Preceding 12 Months
(If changed, publisher must submit explanation of change with this statement.)

10. Extent and Nature of Circulation *(See instructions on reverse side)*	Average No. Copies Each Issue During Preceding 12 Months	Actual No. Copies of Single Issue Published Nearest to Filing Date
A. Total No. Copies *(Net Press Run)*	4845	4737
B. Paid and/or Requested Circulation 1. Sales through dealers and carriers, street vendors and counter sales	----	----
2. Mail Subscription *(Paid and/or requested)*	3961	3908
C. Total Paid and/or Requested Circulation *(Sum of 10B1 and 10B2)*	3961	3908
D. Free Distribution by Mail, Carrier or Other Means Samples, Complimentary, and Other Free Copies	174	178
E. Total Distribution *(Sum of C and D)*	4135	4086
F. Copies Not Distributed 1. Office use, left over, unaccounted, spoiled after printing	710	651
2. Return from News Agents	----	----
G. TOTAL *(Sum of E, F1 and 2—should equal net press run shown in A)*	4845	4737

11. I certify that the statements made by me above are correct and complete	Signature and Title of Editor, Publisher, Business Manager, or Owner
	Michael Kelly, Managing Editor

PS Form 3526, Feb. 1989 (See instructions on reverse)

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

110159

Digitized by Arya Samaj Foundation Chennai and eGangotri

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar